HC70A Winter 2006

Professor Bob Goldberg

Lecture #7

The Human Genome, Human Gene Diversity, + Are there Races?

Themes/Concepts

1. Human Genomes
2. Mitochondrial Genome + Inheritance
3. Human Nuclear Genome
4. SNPs + Origins of Human Gene Diversity/Variation
5. Uses of SNPs
6. Haplotypes / Uses
7. Using Haplotypes to Trace Human Origins
8. Are there Races?
9. Skin Color Genes
10. Genetic Variation within + between Populations
11. "Where" does Most Human Gene Variation Reside?
12. Origins of Race Concept

— Stop 3/9/06

1hr only

HUMAN GENES ARE present in TWO Compartments -- The Nucleus & The Mitochondria

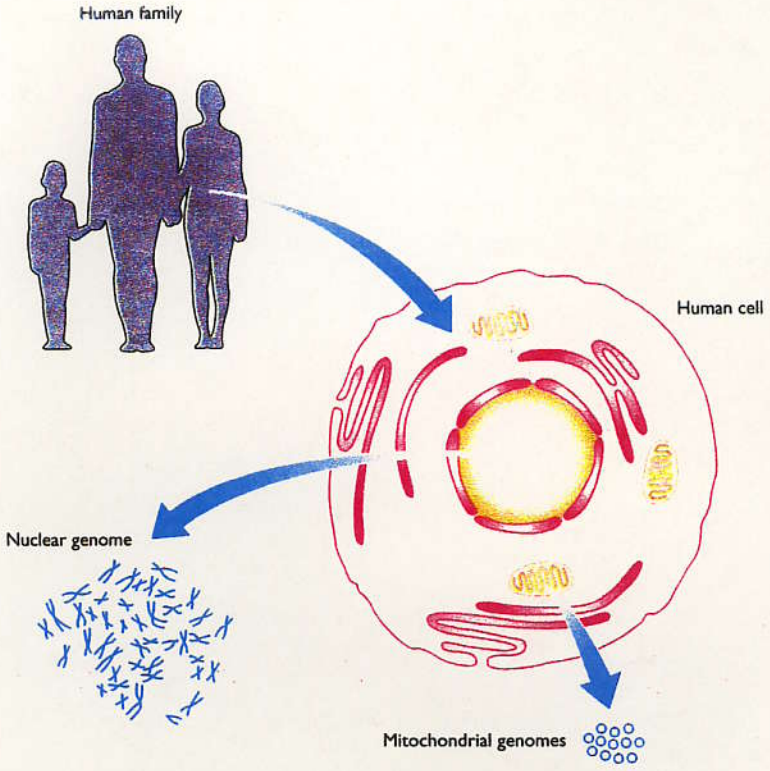


Figure 1.3 The nuclear and mitochondrial components of the human genome.

For more details on the anatomy of the human genome, see Section 6.1.

Genes in BOTH compartments are critical for human development -

Features of the Human Nuclear and Mitochondrial Genomes

Table 9.1: The human nuclear and mitochondrial genomes

	Nuclear genome	Mitochondrial genome
Size	3200 Mb	16.6 kb
No. of different DNA molecules	23 (in XX cells) or 24 (in XY cells); all linear	One circular DNA molecule
Total no. of DNA molecules per cell	46 in diploid cells, but varies according to ploidy	Often several thousands (but variable – see *Box 9.1*)
Associated protein	Several classes of histone and nonhistone protein	Largely free of protein
No. of genes	~ 30 000–35 000	37
Gene density	~ 1/100 kb	1/0.45 kb
Repetitive DNA	Over 50% of genome, see *Figure 9.1*	Very little
Transcription	The great bulk of genes are transcribed individually (*monocistronic transcription units*)	Co-transcription of multiple genes from both the heavy and the light strands (*polycistronic transcription units*)
Introns	Found in most genes	Absent
% of coding DNA	~ 1.5%	~ 93%
Codon usage	See *Figure 1.22*	See *Figure 1.22*
Recombination	At least once for each pair of homologs at meiosis	Not evident
Inheritance	Mendelian for sequences on X and autosomes; paternal for sequences on Y	Exclusively maternal

The Mitochondrial Genome is A SMALL CIRCLE containing only 37 Genes

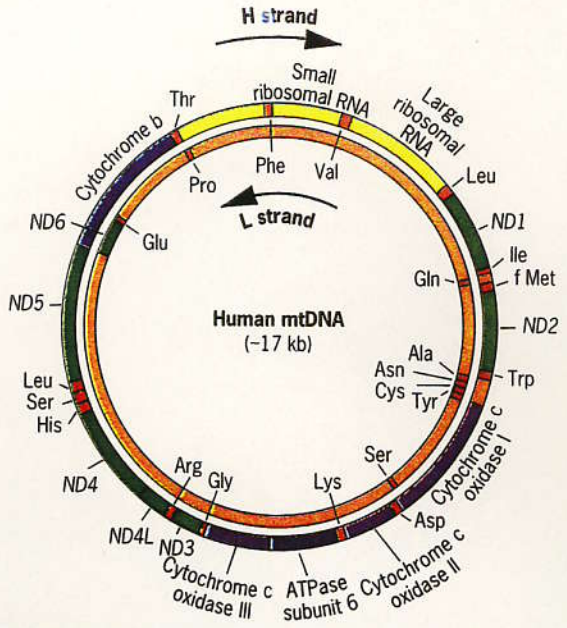
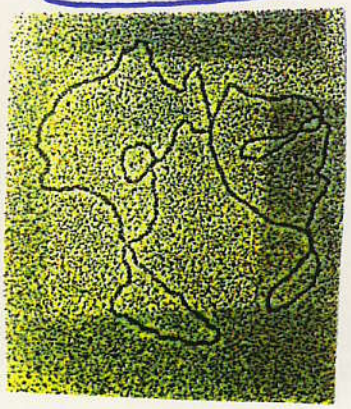
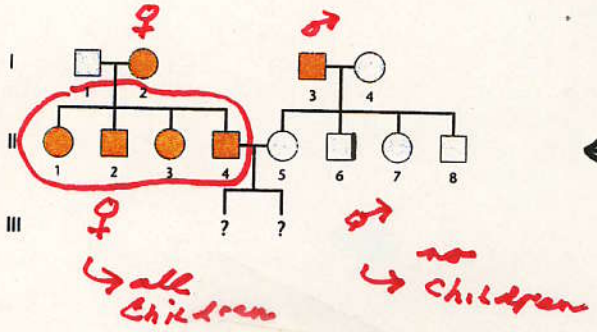


Figure 19.14 Map of human mtDNA showing the pattern of transcription. Genes on the inner circle are transcribed from the L strand of the DNA, whereas genes on the outer circle are transcribed from the H strand of the DNA. Arrows show the direction of transcription. *ND1–6* are genes encoding subunits of the enzyme NADH reductase; the tRNA genes in the mtDNA are indicated by abbreviations for the amino acids.

MtDNA is a CIRCLE



Mitochondrial Genes Are Inherited MATERNALLY



PASSED DIRECTLY FROM Mother to Children

Hypothesis to Explain?

IN A CLONING EXPERIMENT MOST OF MITOCHONDRIA COMES FROM EGG DONOR!

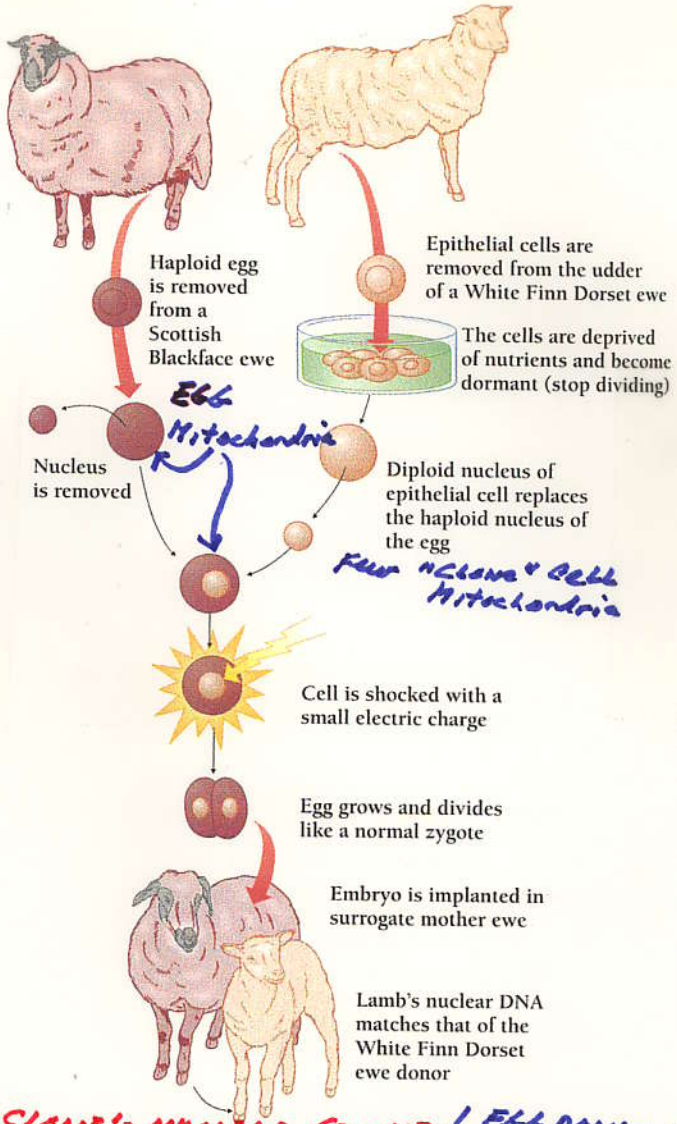


Figure 44-16 **Cloning Dolly.** The trick to cloning Dolly was to make differentiated cells less differentiated. By depriving the cultured udder cells of nutrients, the researchers induced the nuclei to enter a dormant state.

∴ CLONE NOT TECHNICALLY A CLONE WITH RESPECT TO MITOCHONDRIAL GENES

WILL BE MOSAIC - MOST MT FROM EGG DONOR - FEW MT FROM CELL USED FOR CLONE'S GENOME!!

ANOTHER POTENTIAL PROBLEM SOURCE FOR CLONE DEVELOPMENT!

Several Mitochondrial Diseases Occur in Humans

In order for a human disorder to be attributable to genetically altered mitochondria, several criteria must be met.

1. Inheritance must exhibit a maternal rather than a Mendelian pattern.
2. The disorder must reflect a deficiency in the bioenergetic function of the organelle.
3. There must be a specific genetic mutation in one of the mitochondrial genes.

Thus far, several cases are known to demonstrate these characteristics. For example, **myoclonic epilepsy and ragged red fiber disease (MERRF)** demonstrates a pattern of inheritance consistent with maternal inheritance. Only offspring of affected mothers inherit the disorder; the offspring of affected fathers are all normal. Individuals with this rare disorder express deafness, dementia, and seizures. Both muscle fibers and mitochondria are affected; the aberrant mitochondria characterize what are described as ragged red fibers (**RRFs**) of skeletal muscle (Figure 8.5). Analysis of mtDNA has revealed a mutation in one of the mitochondrial genes encoding a transfer RNA. This genetic alteration apparently interferes with translation within the organelle, which in turn leads to the various manifestations of the disorder.

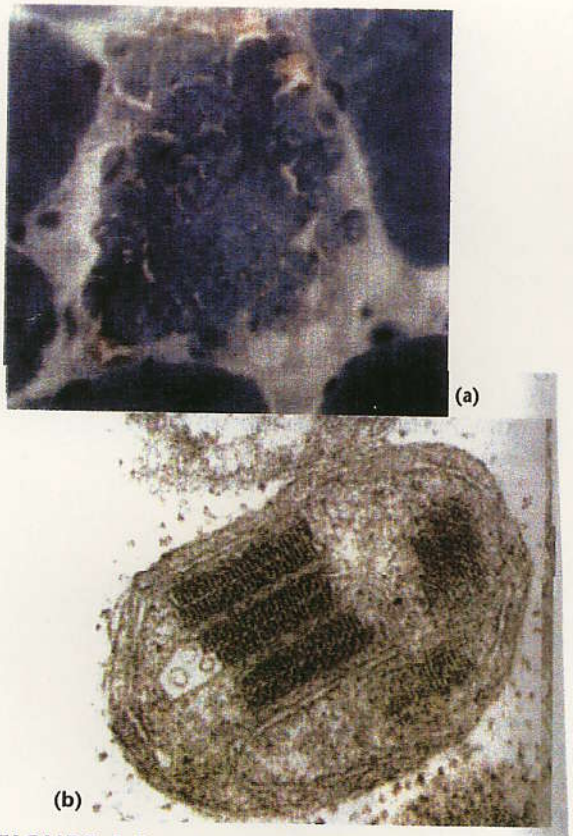


FIGURE 8.5 Mitochondrial myopathy in skeletal muscle cells of a patient with MERFF. Part (a) shows a ragged red fiber with abnormal mitochondria. Part (b) shows an abnormal mitochondrion revealing paracrystalline arrays within it.

A second disorder, **Leber's hereditary optic neuropathy (LHON)**, also exhibits maternal inheritance as well as mtDNA lesions. The disorder is characterized by sudden bilateral blindness. The average age of vision loss is 27, but onset is quite variable. Four mutations have been identified, all of which disrupt normal oxidative phosphorylation. Over 50 percent of cases are due to a mutation at a specific position in the mitochondrial gene encoding a subunit of NADH dehydrogenase so that the amino acid arginine is converted to histidine. This mutation is transmitted to all maternal offspring. It is interesting to note that in many instances of LHON, there is no family history; a significant number of cases appear to result from "new" mutations.

Individuals severely affected by a third disorder, **Kearns-Sayre syndrome (KSS)**, lose their vision, undergo hearing loss, and display heart conditions. The genetic basis of KSS involves deletions at various positions within mtDNA. Many KSS patients are symptom-free as children but display progressive symptoms as adults. The proportion of mtDNAs that reveal deletions increases as the severity of symptoms increases.

The study of hereditary mitochondrial-based disorders provides insights into the importance and genetic basis of this organelle during normal development, as well as the relationship between mitochondrial function and neuromuscular disorders. Such study has also suggested a hypothesis for aging based on the progressive accumulation of mtDNA mutations and the accompanying loss of mitochondrial function.

Mitochondrial Mutations LEADING TO DISEASES

Table 16.1 Phenotypes associated with some mitochondrial mutations

Nucleotide changed	Mitochondrial component affected	Phenotype[a]
3460	ND1 of Complex I[b]	LHON
11778	ND4 of Complex I	LHON
14484	ND6 of Complex I	LHON
8993	ATP6 of Complex V[b]	NARP
3243	$tRNA^{Leu(UUR)}$[c]	MELAS, PEO
3271	$tRNA^{Leu(UUR)}$	MELAS
3291	$tRNA^{Leu(UUR)}$	MELAS
3251	$tRNA^{Leu(UUR)}$	PEO
3256	$tRNA^{Leu(UUR)}$	PEO
5692	$tRNA^{Asn}$	PEO
5703	$tRNA^{Asn}$	PEO, myopathy
5814	$tRNA^{Cys}$	Encephalopathy
8344	$tRNA^{Lys}$	MERRF
8356	$tRNA^{Lys}$	MERRF
9997	$tRNA^{Gly}$	Cardiomyopathy
10006	$tRNA^{Gly}$	PEO
12246	$tRNA^{Ser(AGY)}$[c]	PEO
14709	$tRNA^{Glu}$	Myopathy
15923	$tRNA^{Thr}$	Fatal infantile multisystem disorder
15990	$tRNA^{Pro}$	Myopathy

[a]LHON Leber's hereditary optic neuropathy; NARP Neurogenic muscle weakness, ataxia, retinitis pigmentosa; MERRF Myoclonic epilepsy and ragged-red fiber syndrome; MELAS Mitochondrial myopathy, encephalopathy, lactic acidosis, stroke-like episodes; PEO Progressive external ophthalmoplegia

[b]Complex I is NADH dehydrogenase. Complex V is ATP synthase.

[c]In $tRNA^{Leu(UUR)}$, the R stands for either A or G; in $tRNA^{Ser(AGY)}$, the Y stands for either T or C.

www.mitomap.org

MAP OF MITOCHONDRIAL GENES AND CORRESPONDING DISEASES

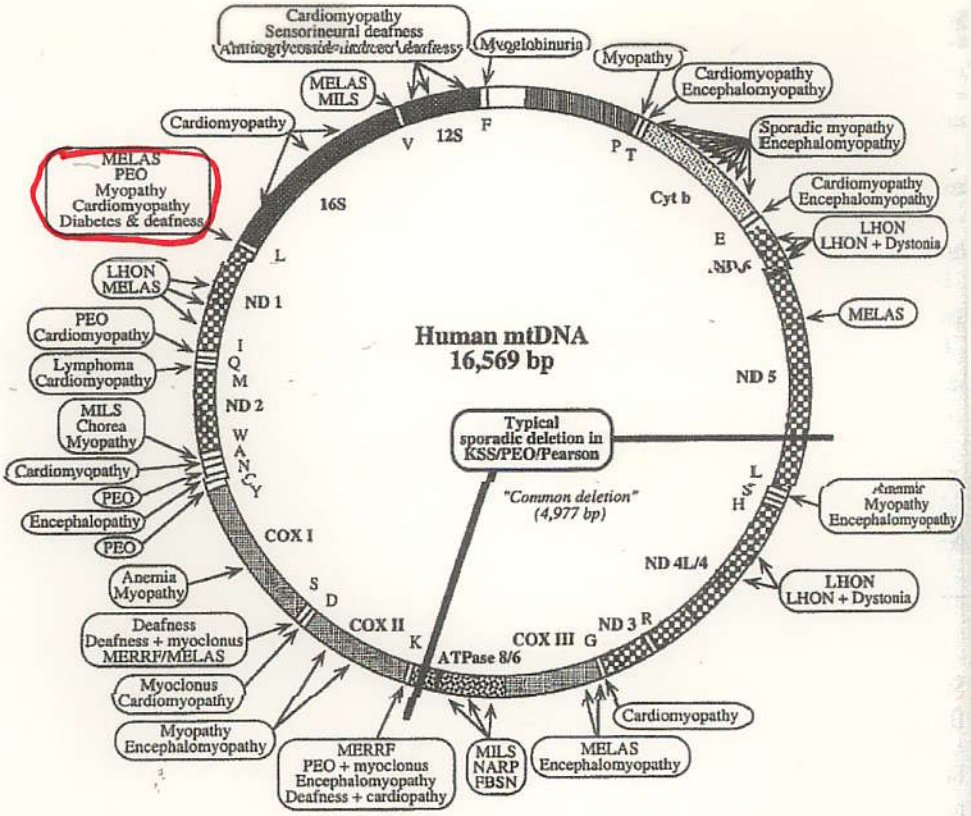


FIGURE 18.16. Morbidity map of the human mitochondrial genome. Abbreviations are for the genes encoding seven subunits of complex I (ND), three subunits of cytochrome c oxidase (COX), cytochrome b (Cyt b), and the two subunits of ATP synthase (ATPase 6 and 8). 12S and 16S refer to ribosomal RNAs; 22 transfer RNAs are identified by the one-letter codes for the corresponding amino acids. FBSN, familial bilateral striatal necrosis; KSS, Kearns–Sayre syndrome; LHON, Leber hereditary optic neuropathy; MELAS, mitochondrial encephalomyopathy, lactic acidosis, and strokelike episodes; MERRF, myoclonic epilepsy with ragged-red fibers; MILS, maternally inherited Leigh syndrome; NARP, neuropathy, ataxia, retinitis pigmentosa; PEO, progressive external ophthalmoplegia. From DiMauro and Schon (2001). Used with permission.

Mitochondrial Diseases ARE Inherited Maternally

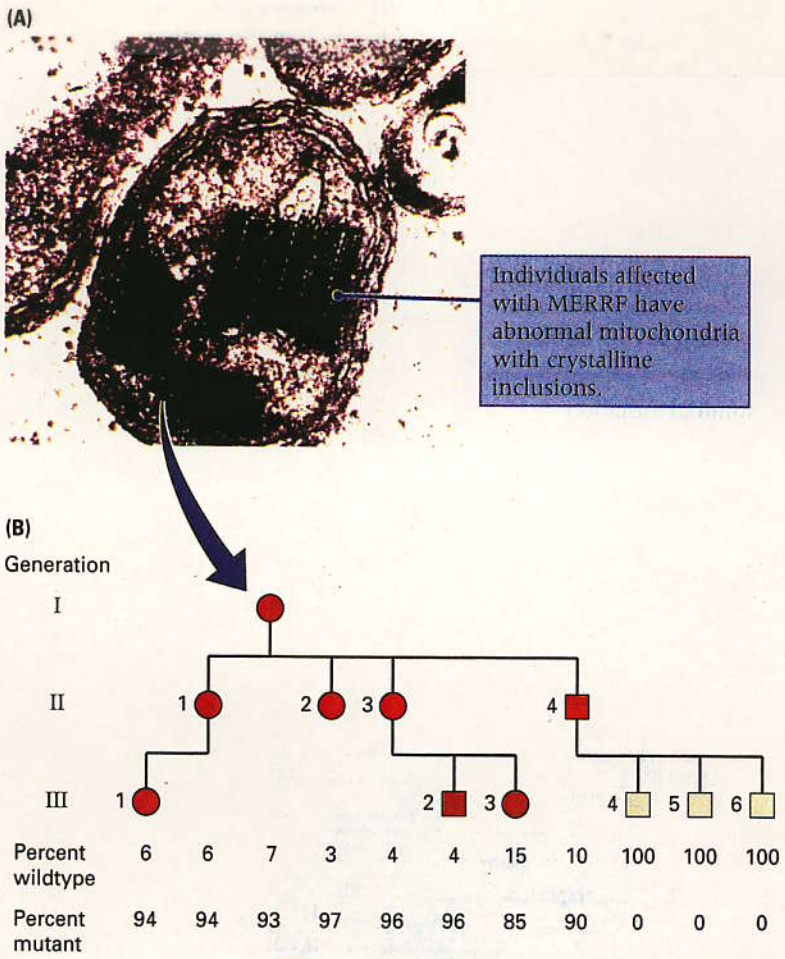


Figure 16.2 Inheritance of myoclonic epilepsy with ragged-red fiber disease (MERRF) in humans. (A) Electron micrograph of an abnormal MERRF mitochondrion containing paracrystalline inclusions. (B)The pedigree shows inheritance of MERRF in one family and the percentage of the mitochondria in each person found to be wildtype or mutant. [Micrograph courtesy of D. C. Wallace, from J. M. Shoffner, M. T. Lott, A. M. S. Lezza, P. Seibel, S. W. Ballinger, and D. C. Wallace. 1990. *Cell* 61: 931.]

Never PASSED ON FROM Diseased ♂ to Children

Mitochondria Absent FROM Sperm "Head"

Maternal Inheritance of Mitochondrial DNA

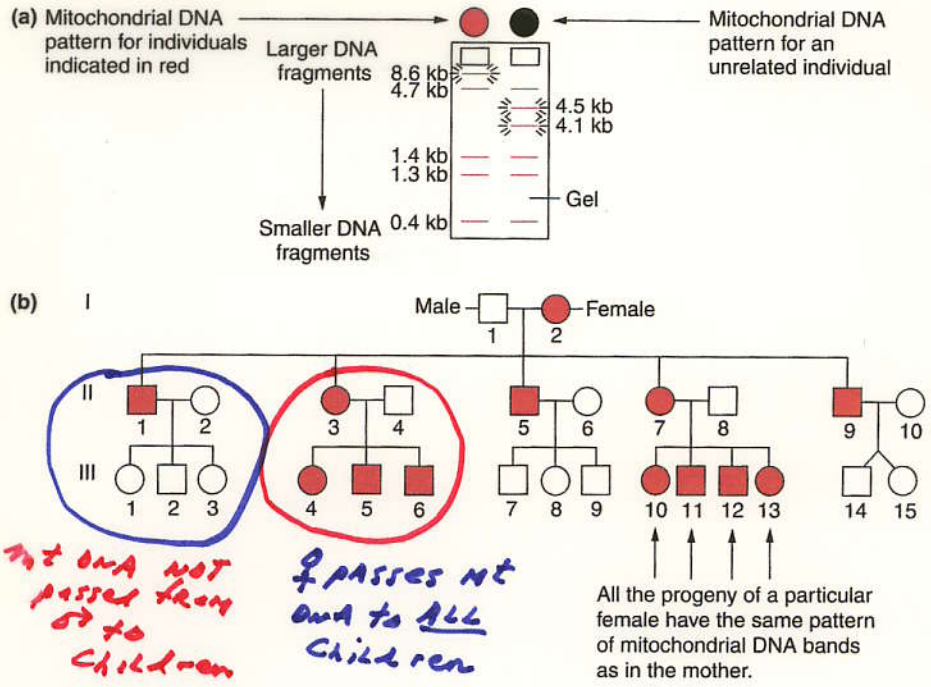
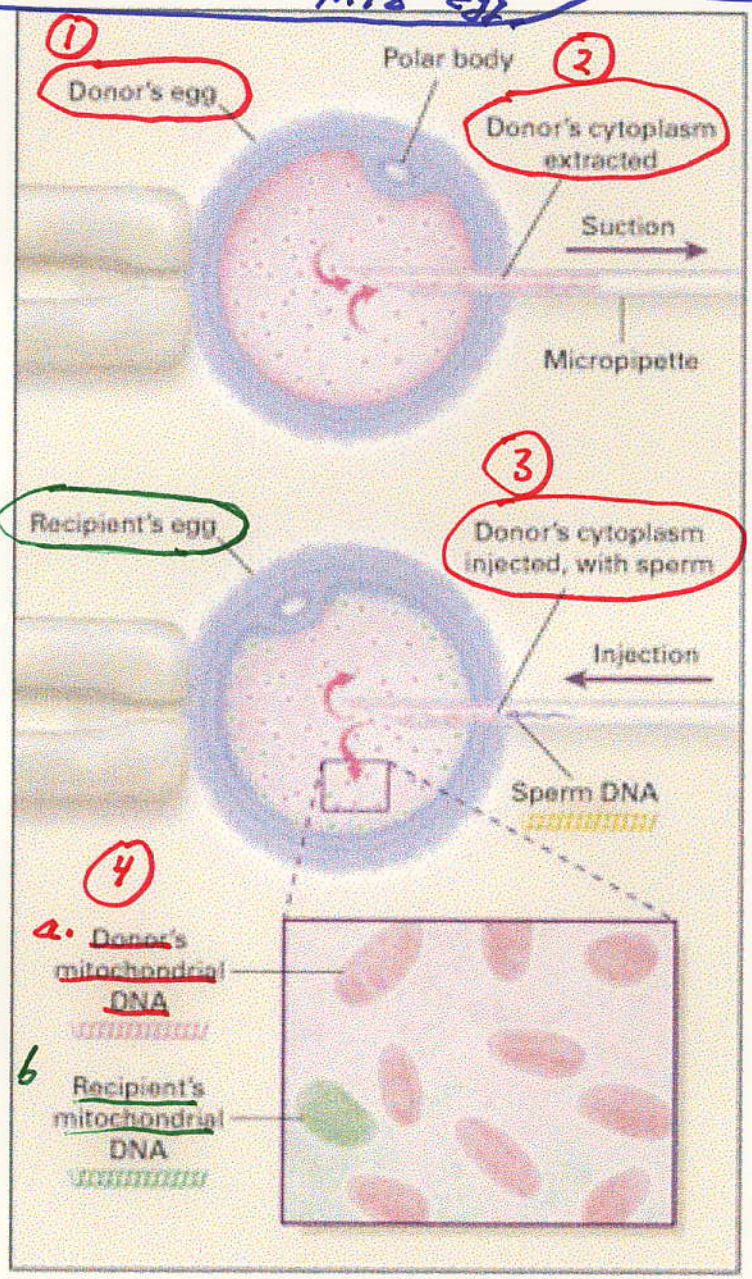


Figure 8.11 Maternal DNA Pattern of Inheritance Key genes for cell respiration and mitochondrial function are located in a small DNA ring in human mitochondria. Because mitochondria are contributed by the egg before fertilization, DNA can be traced through the maternal line with fingerprinting of the mitochondrial DNA.

♂ mt DNA NOT passed from ♂ to children

♀ passes mt DNA to ALL children

OOPLASMIC TRANSFER TO INJECT HEALTHY Mitochondrial Genomes into Egg



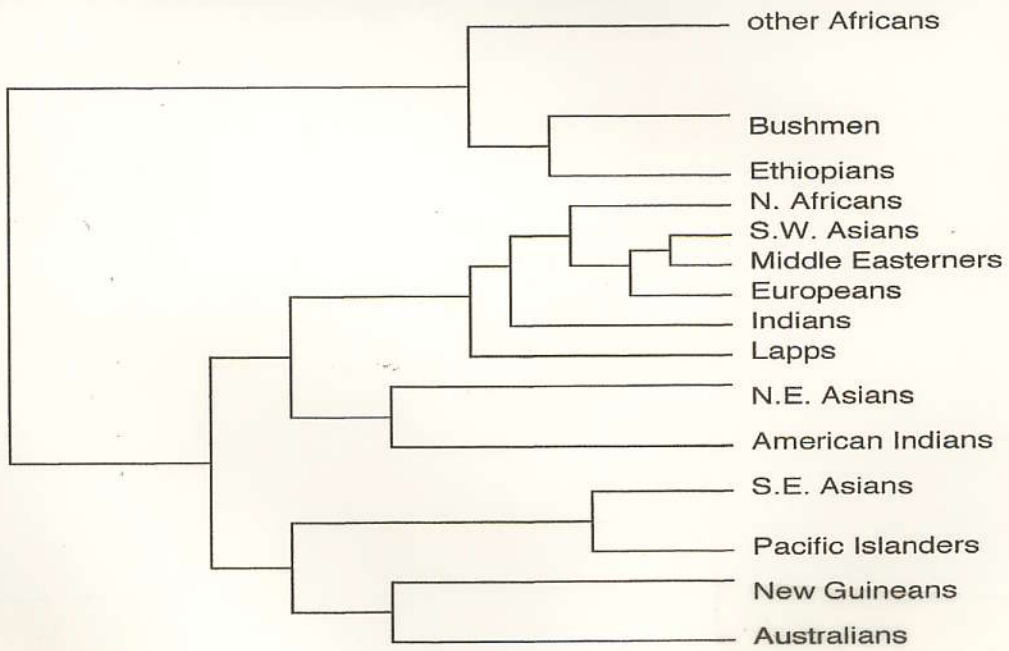
FORM OF Gene Therapy That's Inherited

MOSAIC!

Complements Mt. Gene Detect in Recipient's Egg

USING Mt DNA Polymorphisms to construct trees of Human origins

Evolutionary Tree

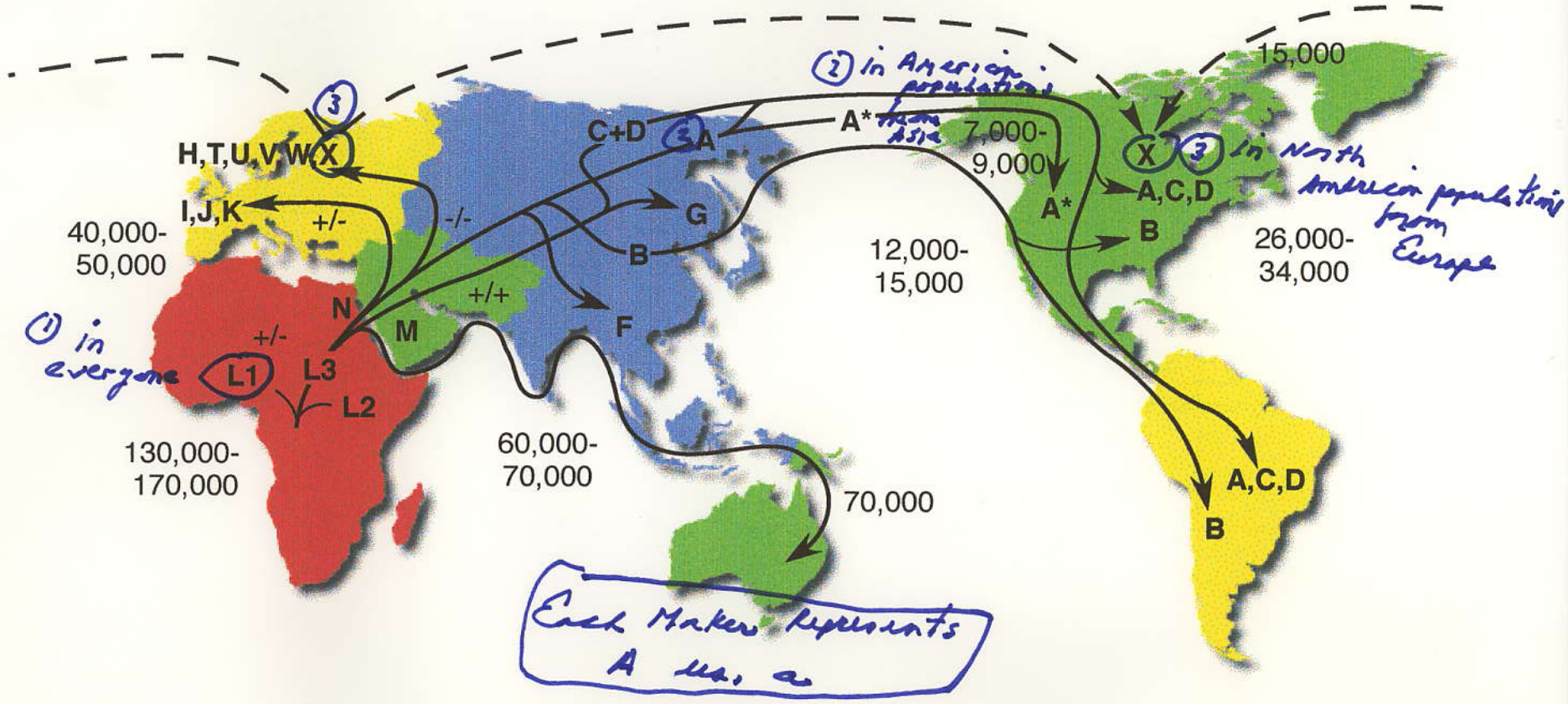


Passed on FROM "EVE" to US!

USING MT DNA RFLPs & SNPs

Human mtDNA Migrations

http://www.mitomap.org/mitomap/WorldMigrations.pdf



+/-, +/+, or -/- = Dde I 10394 / Alu I 10397
* = Rsa I 16329

Mutation rate = 2.2 - 2.9 % / MYR
Time estimates are YBP

(1) ALL populations will contain the oldest/ancient alleles!

(2) New alleles/variants rise in populations as they move/migrate & are fixed in the founder population

THE HUMAN GENOME SEQUENCE

articles

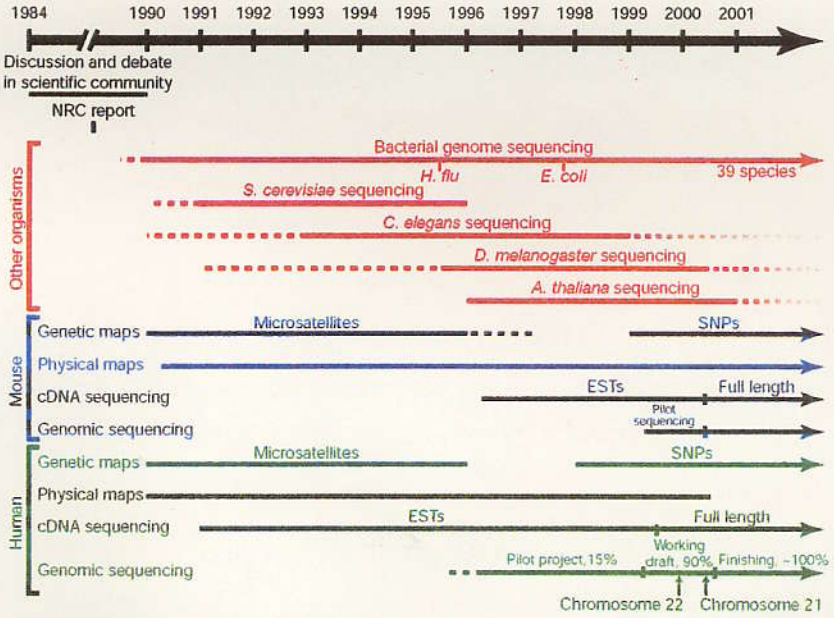
Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium*

* *A partial list of authors appears on the opposite page. Affiliations are listed at the end of the paper.*

The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution. Here we report the results of an international collaboration to produce and make freely available a draft sequence of the human genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequence.

NATURE | VOL 409 | 15 FEBRUARY 2001 | www.nature.com



DONE!

Figure 1 Timeline of large-scale genomic analyses. Shown are selected components of work on several non-vertebrate model organisms (red), the mouse (blue) and the human (green) from 1990; earlier projects are described in the text. SNPs, single nucleotide polymorphisms; ESTs, expressed sequence tags.

13

THE HUMAN GENOME SEQUENCE IS THE RESULT OF AN INTERNATIONAL COLLABORATION

Genome Sequencing Centres (Listed in order of total genomic sequence contributed, with a partial list of personnel. A full list of contributors at each centre is available as Supplementary Information.)

Whitehead Institute for Biomedical Research, Center for Genome Research: Eric S. Lander[1]*, Lauren M. Linton[1], Bruce Birren[1]*, Chad Nusbaum[1]*, Michael C. Zody[1]*, Jennifer Baldwin[1], Keri Devon[1], Ken Dewar[1], Michael Doyle[1], William FitzHugh[1]*, Roel Funke[1], Diane Gage[1], Katrina Harris[1], Andrew Heaford[1], John Howland[1], Lisa Kann[1], Jessica Lehoczky[1], Rosie LeVine[1], Paul McEwan[1], Kevin McKernan[1], James Meldrim[1], Jill P. Mesirov[1]*, Cher Miranda[1], William Morris[1], Jerome Naylor[1], Christina Raymond[1], Mark Rosetti[1], Ralph Santos[1], Andrew Sheridan[1], Carrie Sougnez[1], Nicole Stange-Thomann[1], Nikola Stojanovic[1], Aravind Subramanian[1] & Dudley Wyman[1]

The Sanger Centre: Jane Rogers[2], John Sulston[2]*, Rachael Ainscough[2], Stephan Beck[2], David Bentley[2], John Burton[2], Christopher Clee[2], Nigel Carter[2], Alan Coulson[2], Rebecca Deadman[2], Panos Deloukas[2], Andrew Dunham[2], Ian Dunham[2], Richard Durbin[2]*, Lisa French[2], Darren Grafham[2], Simon Gregory[2], Tim Hubbard[2]*, Sean Humphray[2], Adrienne Hunt[2], Matthew Jones[2], Christine Lloyd[2], Amanda McMurray[2], Lucy Matthews[2], Simon Mercer[2], Sarah Milne[2], James C. Mullikin[2]*, Andrew Mungall[2], Robert Plumb[2], Mark Ross[2], Ratna Shownkeen[2] & Sarah Sims[2]

Washington University Genome Sequencing Center: Robert H. Waterston[3]*, Richard K. Wilson[3], LaDeana W. Hillier[3]*, John D. McPherson[3], Marco A. Marra[3], Elaine R. Mardis[3], Lucinda A. Fulton[3], Asif T. Chinwalla[3]*, Kymberlie H. Pepin[3], Warren R. Gish[3], Stephanie L. Chissoe[3], Michael C. Wendl[3], Kim D. Delehaunty[3], Tracie L. Miner[3], Andrew Delehaunty[3], Jason B. Kramer[3], Lisa L. Cook[3], Robert S. Fulton[3], Douglas L. Johnson[3], Patrick J. Minx[3] & Sandra W. Clifton[3]

US DOE Joint Genome Institute: Trevor Hawkins[4], Elbert Branscomb[4], Paul Predki[4], Paul Richardson[4], Sarah Wenning[4], Tom Slezak[4], Norman Doggett[4], Jan-Fang Cheng[4], Anne Olsen[4], Susan Lucas[4], Christopher Elkin[4], Edward Uberbacher[4] & Marvin Frazier[4]

Baylor College of Medicine Human Genome Sequencing Center: Richard A. Gibbs[5]*, Donna M. Muzny[5], Steven E. Scherer[5], John B. Bouck[5]*, Erica J. Sodergren[5], Kim C. Worley[5]*, Catherine M. Rives[5], James H. Gorrell[5], Michael L. Metzker[5], Susan L. Naylor[6], Raju S. Kucherlapati[7], David L. Nelson, & George M. Weinstock[8]

RIKEN Genomic Sciences Center: Yoshiyuki Sakaki[9], Asao Fujiyama[9], Masahira Hattori[9], Tetsushi Yada[9], Atsushi Toyoda[9], Takehiko Itoh[9], Chiharu Kawagoe[9], Hidemi Watanabe[9], Yasushi Totoki[9] & Todd Taylor[9]

Genoscope and CNRS UMR-8030: Jean Weissenbach[10], Roland Heilig[10], William Saurin[10], Francois Artiguenave[10], Philippe Brottier[10], Thomas Bruls[10], Eric Pelletier[10], Catherine Robert[10] & Patrick Wincker[10]

GTC Sequencing Center: Douglas R. Smith[11], Lynn Doucette-Stamm[11], Marc Rubenfield[11], Keith Weinstock[11], Hong Mei Lee[11] & JoAnn Dubois[11]

Department of Genome Analysis, Institute of Molecular Biotechnology: André Rosenthal[12], Matthias Platzer[12], Gerald Nyakatura[12], Stefan Taudien[12] & Andreas Rump[12]

Beijing Genomics Institute/Human Genome Center: Huanming Yang[13], Jun Yu[13], Jian Wang[13], Guyang Huang[14] & Jun Gu[15]

Multimegabase Sequencing Center, The Institute for Systems Biology: Leroy Hood[16], Lee Rowen[16], Anup Madan[16] & Shizen Qin[16]

Stanford Genome Technology Center: Ronald W. Davis[17], Nancy A. Federspiel[17], A. Pia Abola[17] & Michael J. Proctor[17]

Stanford Human Genome Center: Richard M. Myers[18], Jeremy Schmutz[18], Mark Dickson[18], Jane Grimwood[18] & David R. Cox[18]

University of Washington Genome Center: Maynard V. Olson[19], Rajinder Kaul[19] & Christopher Raymond[19]

Department of Molecular Biology, Keio University School of Medicine: Nobuyoshi Shimizu[20], Kazuhiko Kawasaki[20] & Shinsei Minoshima[20]

University of Texas Southwestern Medical Center at Dallas: Glen A. Evans[21]†, Maria Athanasiou[21] & Roger Schultz[21]

University of Oklahoma's Advanced Center for Genome Technology: Bruce A. Roe[22], Feng Chen[22] & Huaqin Pan[22]

Max Planck Institute for Molecular Genetics: Juliane Ramser[23], Hans Lehrach[23] & Richard Reinhardt[23]

Cold Spring Harbor Laboratory, Lita Annenberg Hazen Genome Center: W. Richard McCombie[24], Melissa de la Bastide[24] & Nellay Dedhia[24]

GBF—German Research Centre for Biotechnology: Helmut Blöcker[25], Klaus Hornischer[25] & Gabriele Nordsiek[25]

*** Genome Analysis Group (listed in alphabetical order, also includes individuals listed under other headings):** Richa Agarwala[26], L. Aravind[26], Jeffrey A. Bailey[27], Alex Bateman[2], Serafim Batzoglou[1], Ewan Birney[28], Peer Bork[29,30], Daniel G. Brown[1], Christopher B. Burge[31], Lorenzo Cerutti[28], Hsiu-Chuan Chen[26], Deanna Church[26], Michele Clamp[2], Richard R. Copley[30], Tobias Doerks[29,30], Sean R. Eddy[32], Evan E. Eichler[27], Terrence S. Furey[33], James Galagan[1], James G. R. Gilbert[2], Cyrus Harmon[34], Yoshihide Hayashizaki[35], David Haussler[36], Henning Hermjakob[28], Karsten Hokamp[37], Wonhee Jang[26], L. Steven Johnson[32], Thomas A. Jones[32], Simon Kasif[38], Arek Kaspryzk[28], Scot Kennedy[39], W. James Kent[40], Paul Kitts[26], Eugene V. Koonin[26], Ian Korf[3], David Kulp[34], Doron Lancet[41], Todd M. Lowe[42], Aoife McLysaght[37], Tarjei Mikkelsen[38], John V. Moran[43], Nicola Mulder[28], Victor J. Pollara[1], Chris P. Ponting[44], Greg Schuler[26], Jörg Schultz[30], Guy Slater[28], Arian F. A. Smit[45], Ella Stupka[26], Joseph Szustakowki[38], Danielle Thierry-Mieg[26], Jean Thierry-Mieg[26], Lukas Wagner[26], John Wallis[3], Raymond Wheeler[34], Alan Williams[34], Yuri I. Wolf[26], Kenneth H. Wolfe[37], Shiaw-Pyng Yang[3] & Ru-Fang Yeh[31]

Scientific management: National Human Genome Research Institute, US National Institutes of Health: Francis Collins[46]*, Mark S. Guyer[46], Jane Peterson[46], Adam Felsenfeld[46]* & Kris A. Wetterstrand[46]; **Office of Science, US Department of Energy:** Aristides Patrinos[47]; **The Wellcome Trust:** Michael J. Morgan[48]

(14)

BUT IT WAS ALSO DONE INDEPENDENTLY BY A COMPANY – CELERA®

The Sequence of the Human Genome

J. Craig Venter,[1]* Mark D. Adams,[1] Eugene W. Myers,[1] Peter W. Li,[1] Richard J. Mural,[1] Granger G. Sutton,[1] Hamilton O. Smith,[1] Mark Yandell,[1] Cheryl A. Evans,[1] Robert A. Holt,[1] Jeannine D. Gocayne,[1] Peter Amanatides,[1] Richard M. Ballew,[1] Daniel H. Huson,[1] Jennifer Russo Wortman,[1] Qing Zhang,[1] Chinnappa D. Kodira,[1] Xiangqun H. Zheng,[1] Lin Chen,[1] Marian Skupski,[1] Gangadharan Subramanian,[1] Paul D. Thomas,[1] Jinghui Zhang,[1] George L. Gabor Miklos,[2] Catherine Nelson,[3] Samuel Broder,[1] Andrew G. Clark,[4] Joe Nadeau,[5] Victor A. McKusick,[6] Norton Zinder,[7] Arnold J. Levine,[7] Richard J. Roberts,[8] Mel Simon,[9] Carolyn Slayman,[10] Michael Hunkapiller,[11] Randall Bolanos,[1] Arthur Delcher,[1] Ian Dew,[1] Daniel Fasulo,[1] Michael Flanigan,[1] Liliana Florea,[1] Aaron Halpern,[1] Sridhar Hannenhalli,[1] Saul Kravitz,[1] Samuel Levy,[1] Clark Mobarry,[1] Knut Reinert,[1] Karin Remington,[1] Jane Abu-Threideh,[1] Ellen Beasley,[1] Kendra Biddick,[1] Vivien Bonazzi,[1] Rhonda Brandon,[1] Michele Cargill,[1] Ishwar Chandramouliswaran,[1] Rosane Charlab,[1] Kabir Chaturvedi,[1] Zuoming Deng,[1] Valentina Di Francesco,[1] Patrick Dunn,[1] Karen Eilbeck,[1] Carlos Evangelista,[1] Andrei E. Gabrielian,[1] Weiniu Gan,[1] Wangmao Ge,[1] Fangcheng Gong,[1] Zhiping Gu,[1] Ping Guan,[1] Thomas J. Heiman,[1] Maureen E. Higgins,[1] Rui-Ru Ji,[1] Zhaoxi Ke,[1] Karen A. Ketchum,[1] Zhongwu Lai,[1] Yiding Lei,[1] Zhenya Li,[1] Jiayin Li,[1] Yong Liang,[1] Xiaoying Lin,[1] Fu Lu,[1] Gennady V. Merkulov,[1] Natalia Milshina,[1] Helen M. Moore,[1] Ashwinikumar K Naik,[1] Vaibhav A. Narayan,[1] Beena Neelam,[1] Deborah Nusskern,[1] Douglas B. Rusch,[1] Steven Salzberg,[12] Wei Shao,[1] Bixiong Shue,[1] Jingtao Sun,[1] Zhen Yuan Wang,[1] Aihui Wang,[1] Xin Wang,[1] Jian Wang,[1] Ming-Hui Wei,[1] Ron Wides,[13] Chunlin Xiao,[1] Chunhua Yan,[1] Alison Yao,[1] Jane Ye,[1] Ming Zhan,[1] Weiqing Zhang,[1] Hongyu Zhang,[1] Qi Zhao,[1] Liansheng Zheng,[1] Fei Zhong,[1] Wenyan Zhong,[1] Shiaoping C. Zhu,[1] Shaying Zhao,[12] Dennis Gilbert,[1] Suzanna Baumhueter,[1] Gene Spier,[1] Christine Carter,[1] Anibal Cravchik,[1] Trevor Woodage,[1] Feroze Ali,[1] Huijin An,[1] Aderonke Awe,[1] Danita Baldwin,[1] Holly Baden,[1] Mary Barnstead,[1] Ian Barrow,[1] Karen Beeson,[1] Dana Busam,[1] Amy Carver,[1] Angela Center,[1] Ming Lai Cheng,[1] Liz Curry,[1] Steve Danaher,[1] Lionel Davenport,[1] Raymond Desilets,[1] Susanne Dietz,[1] Kristina Dodson,[1] Lisa Doup,[1] Steven Ferriera,[1] Neha Garg,[1] Andres Gluecksmann,[1] Brit Hart,[1] Jason Haynes,[1] Charles Haynes,[1] Cheryl Heiner,[1] Suzanne Hladun,[1] Damon Hostin,[1] Jarrett Houck,[1] Timothy Howland,[1] Chinyere Ibegwam,[1] Jeffery Johnson,[1] Francis Kalush,[1] Lesley Kline,[1] Shashi Koduru,[1] Amy Love,[1] Felecia Mann,[1] David May,[1] Steven McCawley,[1] Tina McIntosh,[1] Ivy McMullen,[1] Mee Moy,[1] Linda Moy,[1] Brian Murphy,[1] Keith Nelson,[1] Cynthia Pfannkoch,[1] Eric Pratts,[1] Vinita Puri,[1] Hina Qureshi,[1] Matthew Reardon,[1] Robert Rodriguez,[1] Yu-Hui Rogers,[1] Deanna Romblad,[1] Bob Ruhfel,[1] Richard Scott,[1] Cynthia Sitter,[1] Michelle Smallwood,[1] Erin Stewart,[1] Renee Strong,[1] Ellen Suh,[1] Reginald Thomas,[1] Ni Ni Tint,[1] Sukyee Tse,[1] Claire Vech,[1] Gary Wang,[1] Jeremy Wetter,[1] Sherita Williams,[1] Monica Williams,[1] Sandra Windsor,[1] Emily Winn-Deen,[1] Keriellen Wolfe,[1] Jayshree Zaveri,[1] Karena Zaveri,[1] Josep F. Abril,[14] Roderic Guigó,[14] Michael J. Campbell,[1] Kimmen V. Sjolander,[1] Brian Karlak,[1] Anish Kejariwal,[1] Huaiyu Mi,[1] Betty Lazareva,[1] Thomas Hatton,[1] Apurva Narechania,[1] Karen Diemer,[1] Anushya Muruganujan,[1] Nan Guo,[1] Shinji Sato,[1] Vineet Bafna,[1] Sorin Istrail,[1] Ross Lippert,[1] Russell Schwartz,[1] Brian Walenz,[1] Shibu Yooseph,[1] David Allen,[1] Anand Basu,[1] James Baxendale,[1] Louis Blick,[1] Marcelo Caminha,[1] John Carnes-Stine,[1] Parris Caulk,[1] Yen-Hui Chiang,[1] My Coyne,[1] Carl Dahlke,[1] Anne Deslattes Mays,[1] Maria Dombroski,[1] Michael Donnelly,[1] Dale Ely,[1] Shiva Esparham,[1] Carl Fosler,[1] Harold Gire,[1] Stephen Glanowski,[1] Kenneth Glasser,[1] Anna Glodek,[1] Mark Gorokhov,[1] Ken Graham,[1] Barry Gropman,[1] Michael Harris,[1] Jeremy Heil,[1] Scott Henderson,[1] Jeffrey Hoover,[1] Donald Jennings,[1] Catherine Jordan,[1] James Jordan,[1] John Kasha,[1] Leonid Kagan,[1] Cheryl Kraft,[1] Alexander Levitsky,[1] Mark Lewis,[1] Xiangjun Liu,[1] John Lopez,[1] Daniel Ma,[1] William Majoros,[1] Joe McDaniel,[1] Sean Murphy,[1] Matthew Newman,[1] Trung Nguyen,[1] Ngoc Nguyen,[1] Marc Nodell,[1] Sue Pan,[1] Jim Peck,[1] Marshall Peterson,[1] William Rowe,[1] Robert Sanders,[1] John Scott,[1] Michael Simpson,[1] Thomas Smith,[1] Arlan Sprague,[1] Timothy Stockwell,[1] Russell Turner,[1] Eli Venter,[1] Mei Wang,[1] Meiyuan Wen,[1] David Wu,[1] Mitchell Wu,[1] Ashley Xia,[1] Ali Zandieh,[1] Xiaohong Zhu[1]

(15)

AND COMPLETED IN ONLY Nine Months!

IT WAS A RACE!

Celera Effort

5 people's DNA combined + sequenced

3♀ 2♂

2 Caucasian
1 African American
1 Chinese
1 Hispanic

→ SNPs re allele variability

A 2.91-billion base pair (bp) consensus sequence of the euchromatic portion of the human genome was generated by the whole-genome shotgun sequencing method. The 14.8-billion bp DNA sequence was generated over 9 months from 27,271,853 high-quality sequence reads (5.11-fold coverage of the genome) from both ends of plasmid clones made from the DNA of five individuals. Two assembly strategies—a whole-genome assembly and a regional chromosome assembly—were used, each combining sequence data from Celera and the publicly funded genome effort. The public data were shredded into 550-bp segments to create a 2.9-fold coverage of those genome regions that had been sequenced, without including biases inherent in the cloning and assembly procedure used by the publicly funded group. This brought the effective coverage in the assemblies to eightfold, reducing the number and size of gaps in the final assembly over what would be obtained with 5.11-fold coverage. The two assembly strategies yielded very similar results that largely agree with independent mapping data. The assemblies effectively cover the euchromatic regions of the human chromosomes. More than 90% of the genome is in scaffold assemblies of 100,000 bp or more, and 25% of the genome is in scaffolds of 10 million bp or larger. Analysis of the genome sequence revealed 26,588 protein-encoding transcripts for which there was strong corroborating evidence and an additional ~12,000 computationally derived genes with mouse matches or other weak supporting evidence. Although gene-dense clusters are obvious, almost half the genes are dispersed in low G+C sequence separated by large tracts of apparently noncoding sequence. Only 1.1% of the genome is spanned by exons, whereas 24% is in introns, with 75% of the genome being intergenic DNA. Duplications of segmental blocks, ranging in size up to chromosomal lengths, are abundant throughout the genome and reveal a complex evolutionary history. Comparative genomic analysis indicates vertebrate expansions of genes associated with neuronal function, with tissue-specific developmental regulation, and with the hemostasis and immune systems. DNA sequence comparisons between the consensus sequence and publicly funded genome data provided locations of 2.1 million single-nucleotide polymorphisms (SNPs). A random pair of human haploid genomes differed at a rate of 1 bp per 1250 on average, but there was marked heterogeneity in the level of polymorphism across the genome. Less than 1% of all SNPs resulted in variation in proteins, but the task of determining which SNPs have functional consequences remains an open challenge.

1% Exons
24% Introns
75% Intergenic DNA
100%

2005
5.8 × 10⁶ SNPs
or
1/~600 bp
on average between
2 people

1 SNP/1250 bp on average between two people

Now one SNP/500bp – HAPMAP PROJECT

BUT The International public Sequence is MORE complete than the private one That contains Gaps

READ –

"The GENOME WAR"

by James Shreeve

ISBN 0375406298

JANUARY, 2004

Private vs. Public Genome Projects!

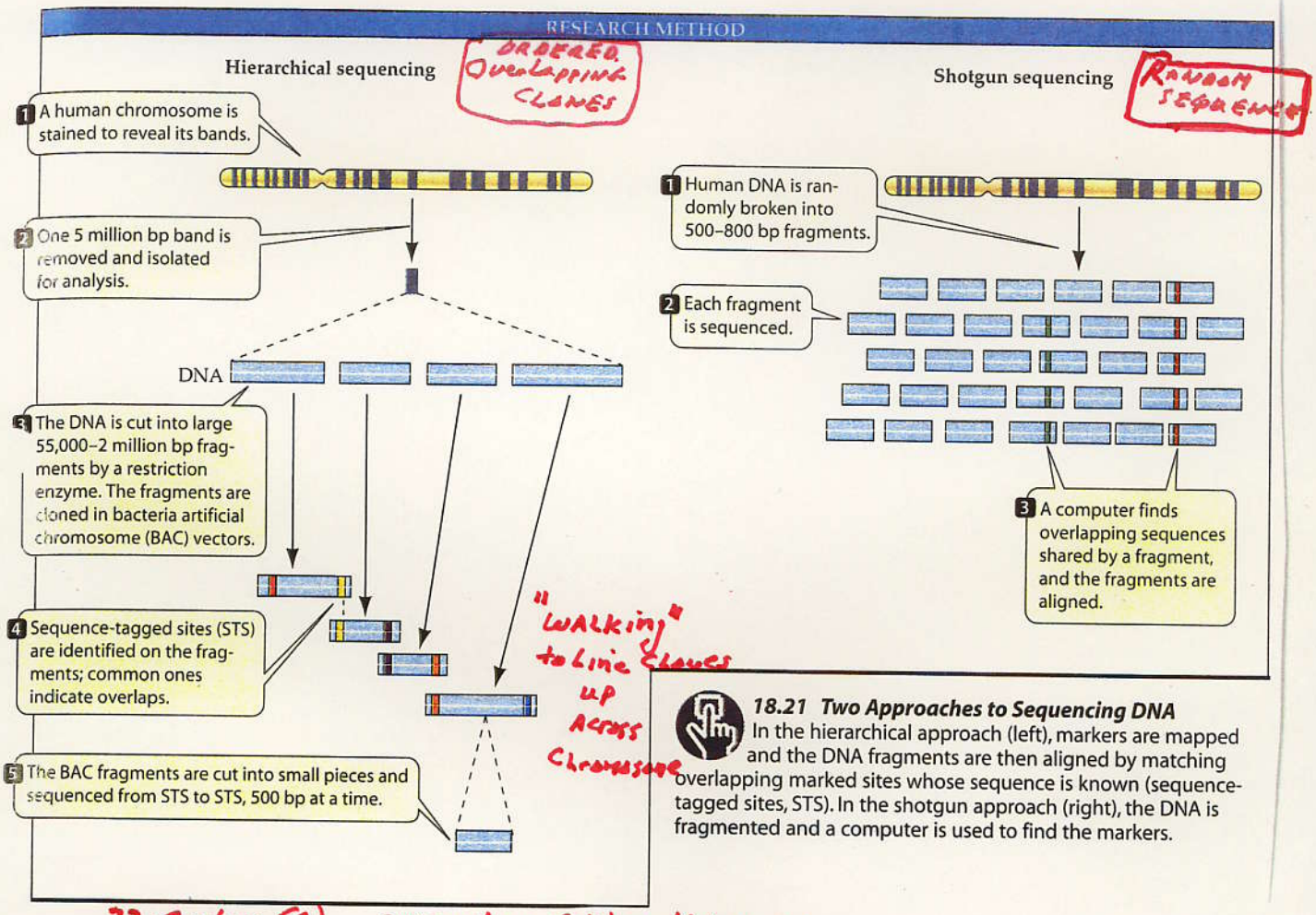
HOW WAS THE HUMAN GENOME SEQUENCED?

TOP DOWN

Physical Map → Sequence

BOTTOM UP

Sequence → Assemble



18.21 Two Approaches to Sequencing DNA In the hierarchical approach (left), markers are mapped and the DNA fragments are then aligned by matching overlapping marked sites whose sequence is known (sequence-tagged sites, STS). In the shotgun approach (right), the DNA is fragmented and a computer is used to find the markers.

23 contigs (♀) 24 contigs (♂) – Y chromosome

PUBLIC EFFORT

CHROMOSOME WALKS → SEQUENCE

BEST
MOST COMPLETE
ENTIRE
SLOW

PRIVATE EFFORT

SHOTGUN

FAST
MANY GAPS
SKELETON

NEEDS PUBLIC DATA TO ASSEMBLE → CHAOS

Approximately Half of the Genes in the Human Genome have Unknown Functions

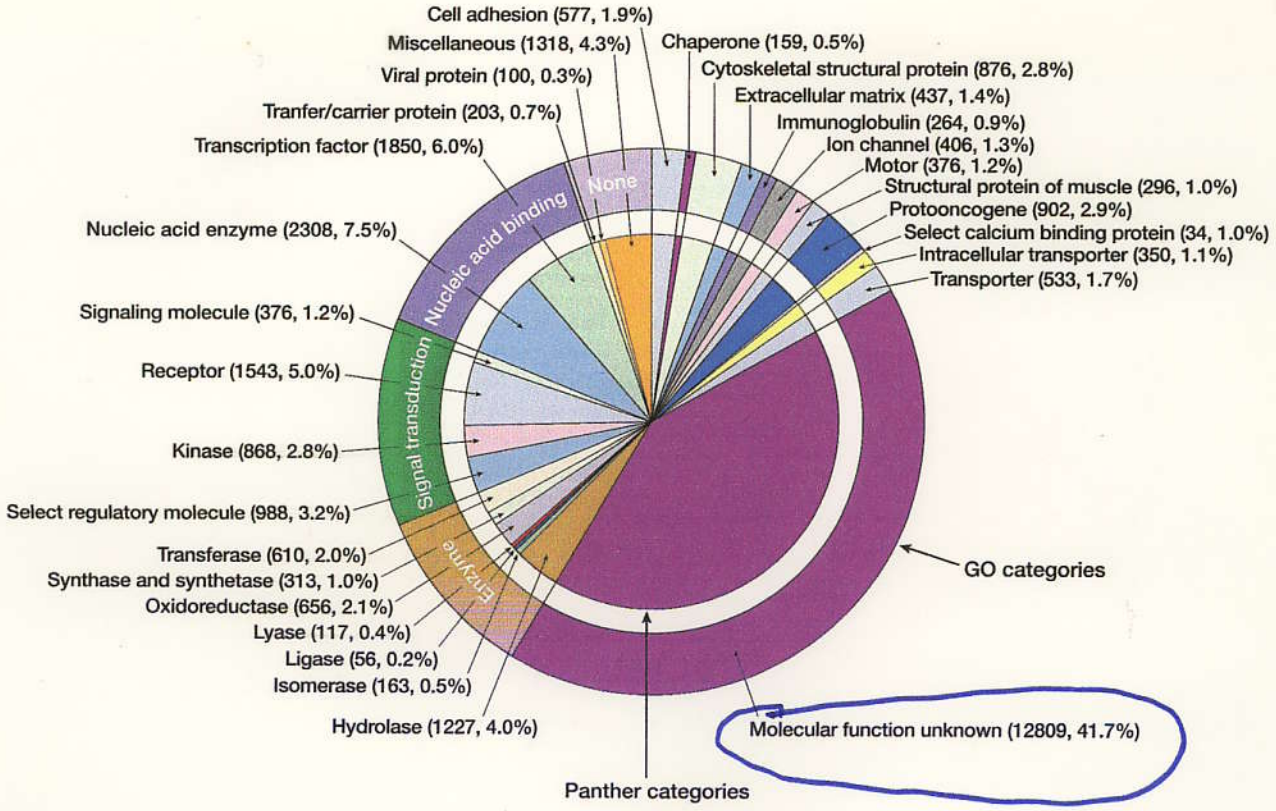


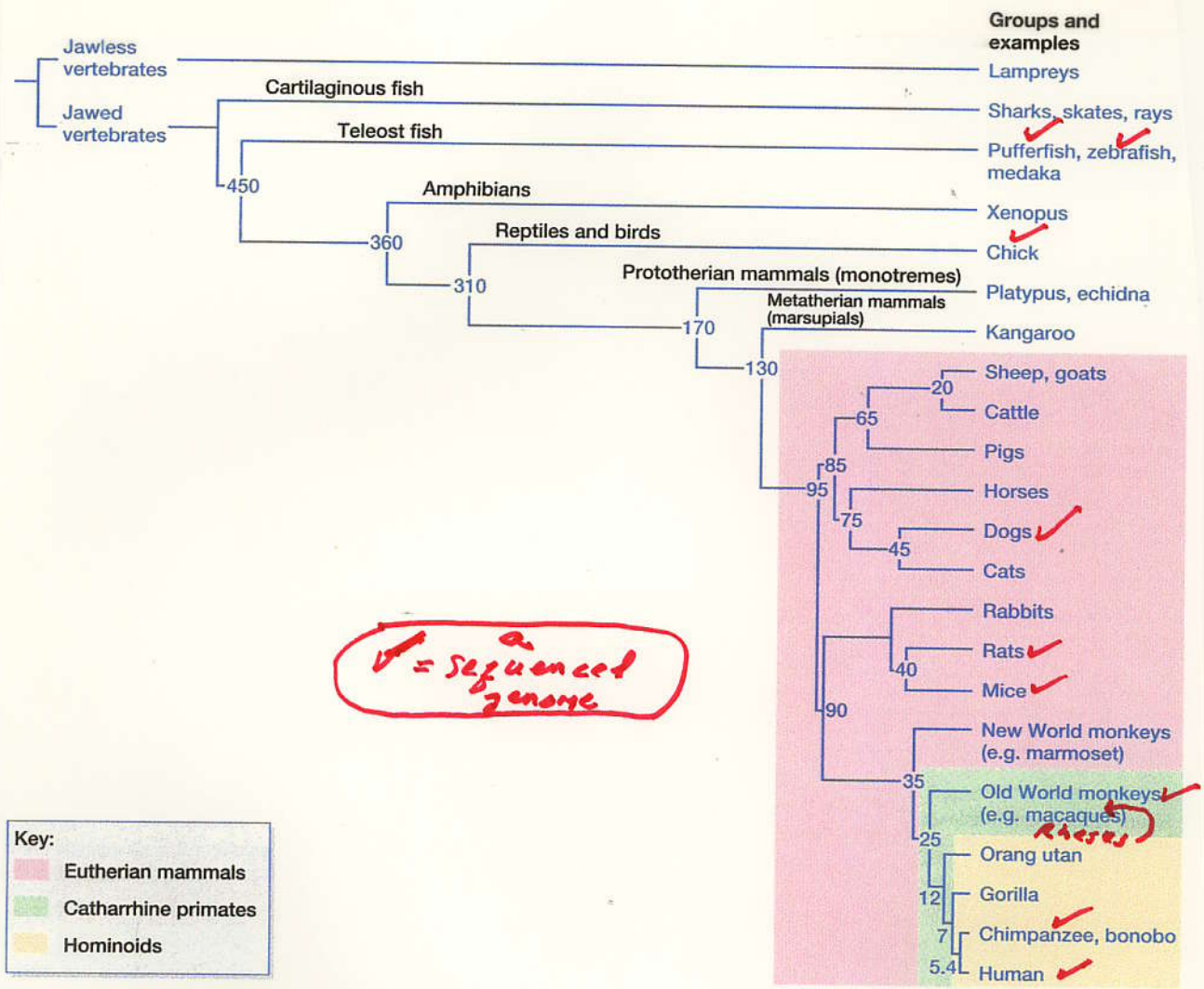
Figure 12.20: A preliminary functional classification of human polypeptide-encoding genes.

Known or predicted functions for 26 383 human polypeptide-encoding genes. Classification is according to the GO molecular function categories as shown in the outer circle (Gene Ontology classification – see Section 8.3.6) or to Celera's Panther molecular function categories (inner circle). Reproduced from Venter *et al.* (2001) *Science* **291**, 1304–1351, with permission from the American Association for the Advancement of Science.

LOTS OF WORK Yet to do!

Use Mouse as Model to Identify Unknown Genes functions!!

VERTABRATE AND MAMMALIAN RELATIONSHIPS



✓ = Sequenced genome

Figure 12.24: A simplified vertebrate phylogeny.

Numbers at nodes show estimated divergence times in millions of years.

COMPARATIVE GENOMICS

HUMAN & CHIMP GENOMES ARE VIRTUALLY IDENTICAL

Chimp Genome Draft Online

The relationship between humans and chimps just got a little easier to understand. This week, the consortium that has been unraveling the DNA sequence of our closest cousin for the past year put its results into the public domain. Robert Waterston of the University of Washington, Seattle, and his colleagues at the Broad Institute in Cambridge, Massachusetts (including the former genome center of the Whitehead Institute for Biomedical Research), and at the Washington University Genome Sequencing Center in St. Louis, Missouri, have determined the order of many of the 3.1 billion bases of a single male chimp's genome. Until now, only pieces of deciphered DNA were available, and their placement along the two dozen chromosomes was uncertain.

On average, each base was sequenced only four times. That's far short of the current 10-fold coverage of the human genome, but it's enough to put together a rough draft with many bases in the right order, which the researchers deposited online in GenBank. The consortium matched up the human and chimp genomes base by base as much as possible, an alignment that will make it easier for researchers to find elusive genes and regulatory regions. The matchup will also highlight specific differences between the two genomes, perhaps further hinting at what set us apart. The sequence is more than researchers could have hoped for a few years ago, says Ajit Varki of the University of California, San Diego, and it "will be most useful" for geneticists trying to find genes responsible for inherited diseases.

Work on the chimp sequence continues, but in the meantime, the consortium is taking the next few months to analyze the data it has. It expects to publish results early next year. –E.P.

CREDITS: (LEFT TO RIGHT) PAT POWERS AND CHERRYL SCHAFER/GETTY IMAGES; C

10,000,000 YEAR Divergence

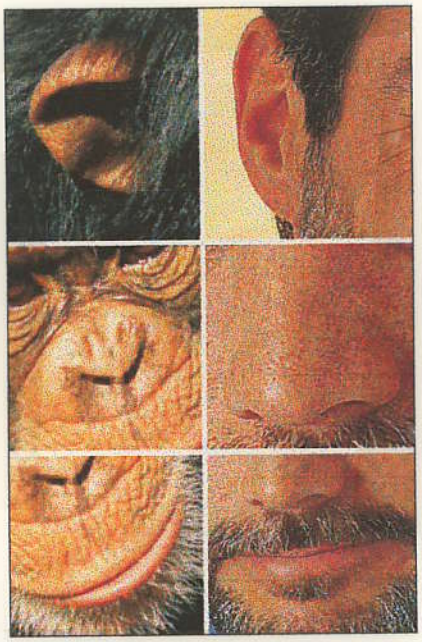
But differences should indicate why a man is a man & a chimp a chimp

key to understanding unique Human Features at Molecular Level!!

Compare all Mammalian Genomes!

EVOLUTION

Genome Comparisons Hold Clues to Human Evolution



Hear no evil. Changes in genes for hearing, olfaction, and speech helped prompt human evolution.

SINGLE BASE PAIR CHANGES (SNPs) ARE FREQUENT IN GENOME

TABLE 9.1 Five Classes of DNA Polymorphism

Class	Cause	Rate of Mutation per Locus per Gamete	Frequency in Genome	Number per Human Genome (on average)
Single base	Mutagens or replication errors	10^{-8}–10^{-9}	1/700 bp	3 million
Microsatellite	Slippage during replication	10^{-3}	1/30,000 bp	100,000
Minisatellite	Unequal crossovers	10^{-3}	Unknown; discovered by chance	Fewer than 100 families known, yielding 1000 copies in all
Deletions	Mutagens; unequal crossovers	Extremely rare	Very low	0 – a few
Duplications	Mutagens; unequal crossovers	Extremely rare	Very low	0 – a few
Other insertions (excluding those resulting from micro- or minisatellite recombination)	Transposable elements	Extremely rare	Very low	0 – a few
Complex haplotype (any locus of 5 kb or more)	Any of the above	Combination of the above	Not applicable	Not applicable

DETECTED USING RESTRICTION ENZYMES (THE OLD FASHIONED WAY) OR BY DIRECT DNA SEQUENCING OF 2 INDIVIDUALS' GENES / GENOMES

SNPs or Single Nucleotide Polymorphisms occur Randomly in the Human Genome

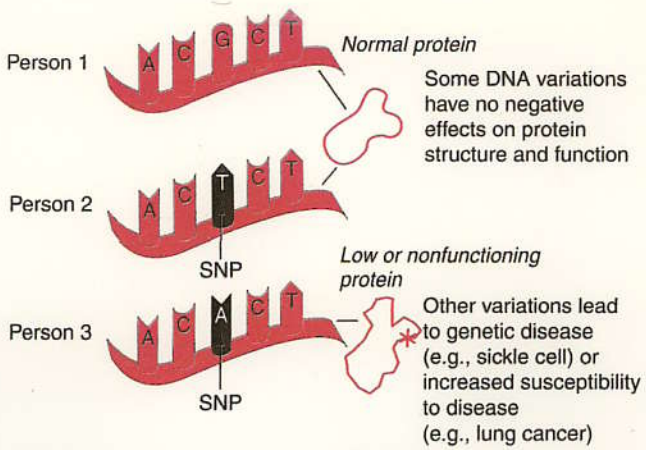


Figure 1.10 Single Nucleotide Polymorphisms A small piece of a gene sequence for three different individuals is represented. For simplicity, only one strand of a DNA molecule is shown. Notice how person 2 has a SNP in this gene which has no effect on protein structure and function. Person 3 however, has a different SNP in the same gene. This subtle genetic change may affect how this person responds to a medical drug or influence the likelihood that person 3 will develop a genetic disease.

used to individualize genomes & associate with disease genes (i.e., disease gene markers)

detected directly by sequencing

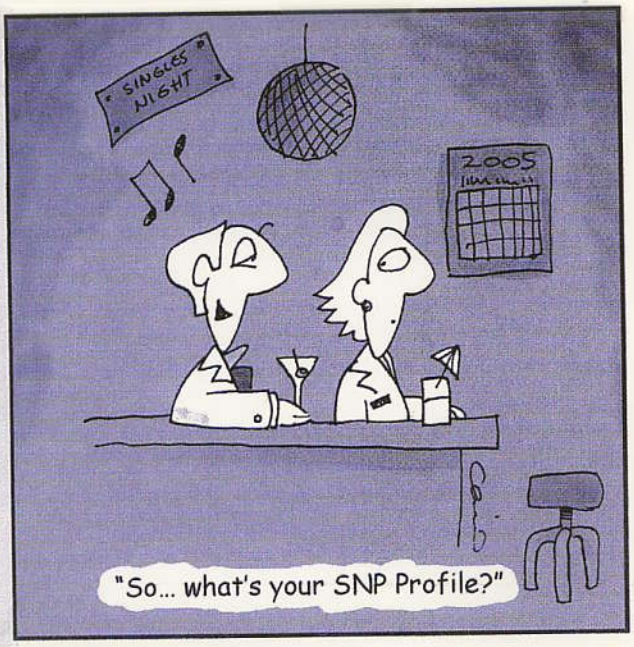


Figure 1.11 Secrets of the Human Genome In the future, we will have unprecedented knowledge of our genetic make-up including SNPs and other markers of genetic diseases. Can you think of possible ethical, legal, and social implications of such information?

LOCATION of SNPs Relative to Genes

TABLE 18-11 dbSNP Statistics (NCBI Genome Build 30, November 2002)

See http://www.ncbi.nlm.nih.gov/SNP/snp_summary.cgi. SNP count is the number of distinct RefSNPs having the noted functional relationship to at least one mRNA in the current assembly. Gene count is the number of distinct locus_id(s) having at least one variation of the noted functional class. (Genes with multiple variations may be counted in multiple classes.)

Functional Classification	SNP Count	Gene Count
Locus region	291,459	26,482
Allele synonymous to contig nucleotide	12,322	7,147
Allele nonsynonymous to contig nucleotide	16,251	8,496
Untranslated region	131,987	13,208
Intron	904,573	22,113
Splice site	277	268
Allele is same as contig nucleotide	28,491	11,621
Coding: synonymy unknown	13,501	3,584

TABLE 18-10 SNP Resources

Resource	Comment	URL
dbSNP	At NCBI	www.ncbi.nlm.nih.gov/SNP
Human *SNP* database	At the Whitehead Institute	http://www-genome.wi.mit.edu/snp/human/
The SNP Consortium (TSC)	A collaboration of industrial and academic laboratories	http://snp.cshl.org

MOST SNPs ARE NOT IN CODING SEQUENCES & HAVE NO PHENOTYPIC EFFECTS!

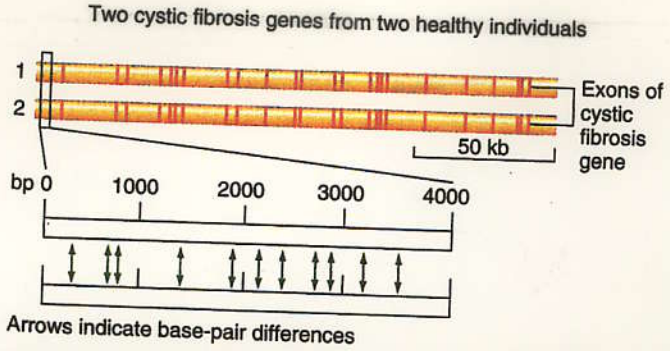


Figure 9.2 Base-pair differences between DNA cloned from the cystic fibrosis locus of two healthy individuals. These base-pair differences have no phenotypic effect; apparently they neither encode nor regulate expressed regions of the gene.

NOTE Differences Between Two Healthy Cystic Fibrosis Genes!

~ 6 x 10⁶ SNPs (2006) :

3.3 x 10⁹ / 6 x 10⁶ = 500 - 600 bp / SNP on average

EACH OF US DIFFERS BY 6 x 10⁶ bp or ~ 0.2% of Genome!!

MOST in NON-CODING/INTRON & Intergenic Regions & Useful As Markers for Forensics, Disease Genes, Populations

USING SNPs as GENE MARKERS

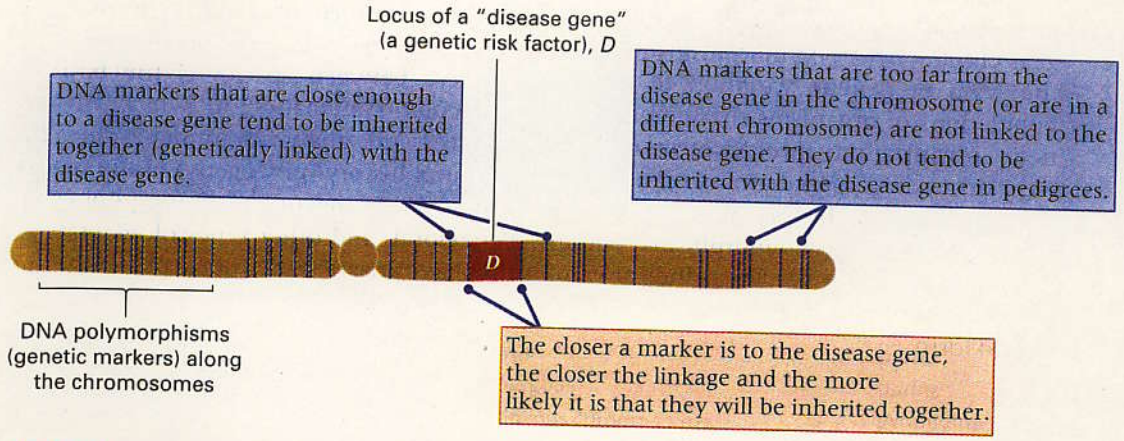


Figure 2.29 Concepts in genetic localization of genetic risk factors for disease. Polymorphic DNA markers (indicated by the vertical lines) that are close to a genetic risk factor (*D*) in the chromosome tend to be inherited together with the disease itself. The genomic location of the risk factor is determined by examining the known genomic locations of the DNA polymorphisms that are linked with it.

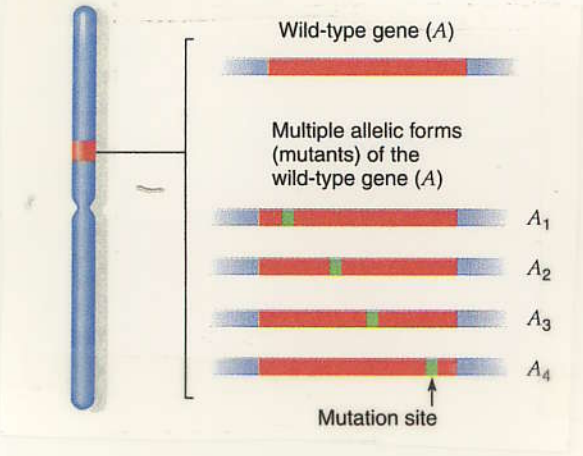
MAP GENES FOR DISEASE SUSCEPTIBILITY – GENES ENCODING COMPLEX (MULTI-GENIC) TRAITS (e.g., Heart Disease, Depression, DRUG Sensitivity, Obesity)

SNPs = INDIVIDUAL GENE PROFILE

↳ INDIVIDUAL Medicine!

SNPs Generate Multiple Alleles in a Population

Figure 4.1
Allelic forms of a gene.



Recall – 2 alleles / individual
May in POPULATIONS!

USES OF SNPs

1. Gene Identity / Allele Marker
2. Disease Gene Identity / Pedigrees / Testing
3. Group Identity / Population History / (Human origins)
4. Individual Identity
   - → Pharmacogenomics
   - → Disease Prevention / Preventative Medicine
   - → DNA Fingerprinting
5. Group Susceptibility to Disease / Drugs
   - → Identify Genes Involved
   - → "Resistance" Genes

Using SNPs to Personalize Medicine
Pharmacogenomics

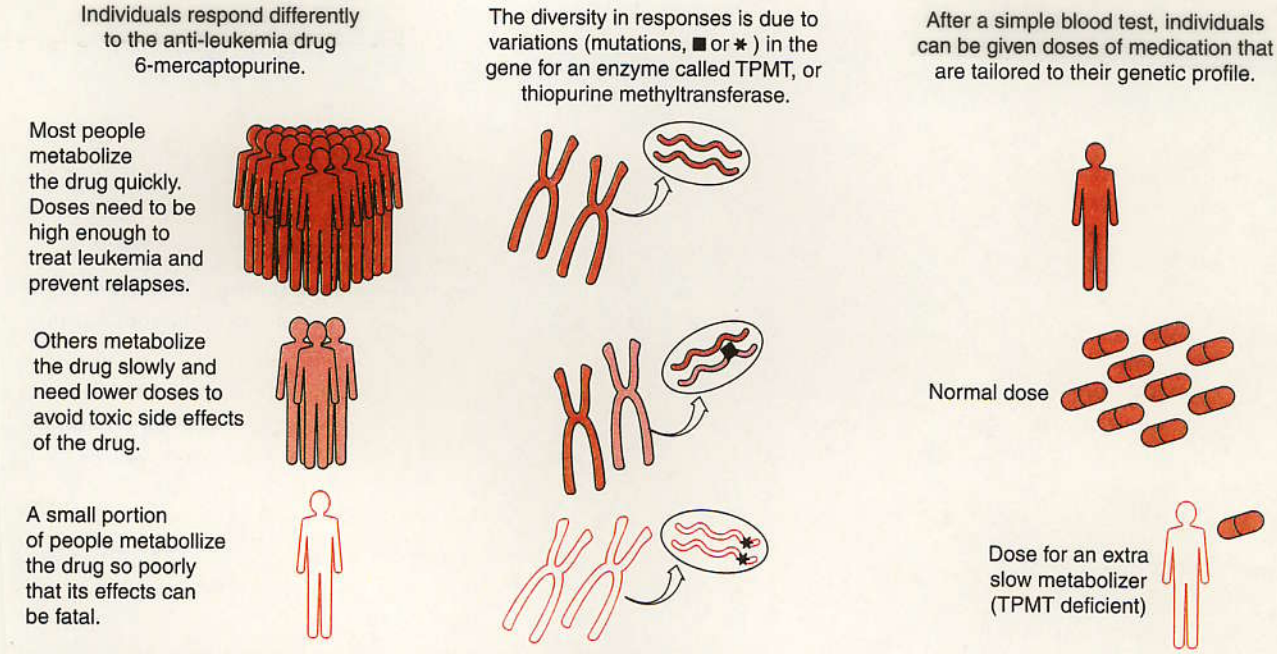
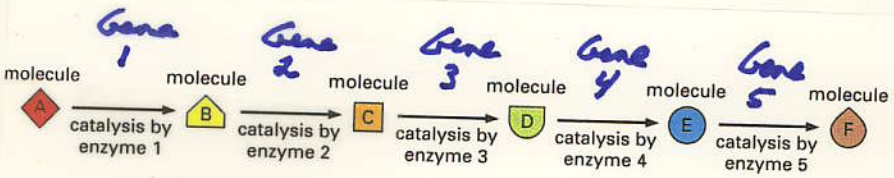


Figure 11.7 Pharmacogenomics Different individuals with the same disease often respond differently to a drug treatment because of subtle differences in gene expression. The dose that works for one person may be toxic for another—this is a basic problem of conventional medicine. Pharmacogenomics holds the promise of customizing medical treatment by determining the appropriate dosage for each individual based on the genes that person expresses.

Using SNPs in Pharmacogenetics



ABBREVIATED AS

Figure 2–34 How a set of enzyme-catalyzed reactions generates a metabolic pathway. Each enzyme catalyzes a particular chemical reaction, leaving the enzyme unchanged. In this example, a set of enzymes acting in series converts molecule A to molecule F, forming a metabolic pathway.

Different alleles can give rise to enzymes that differ slightly in activity

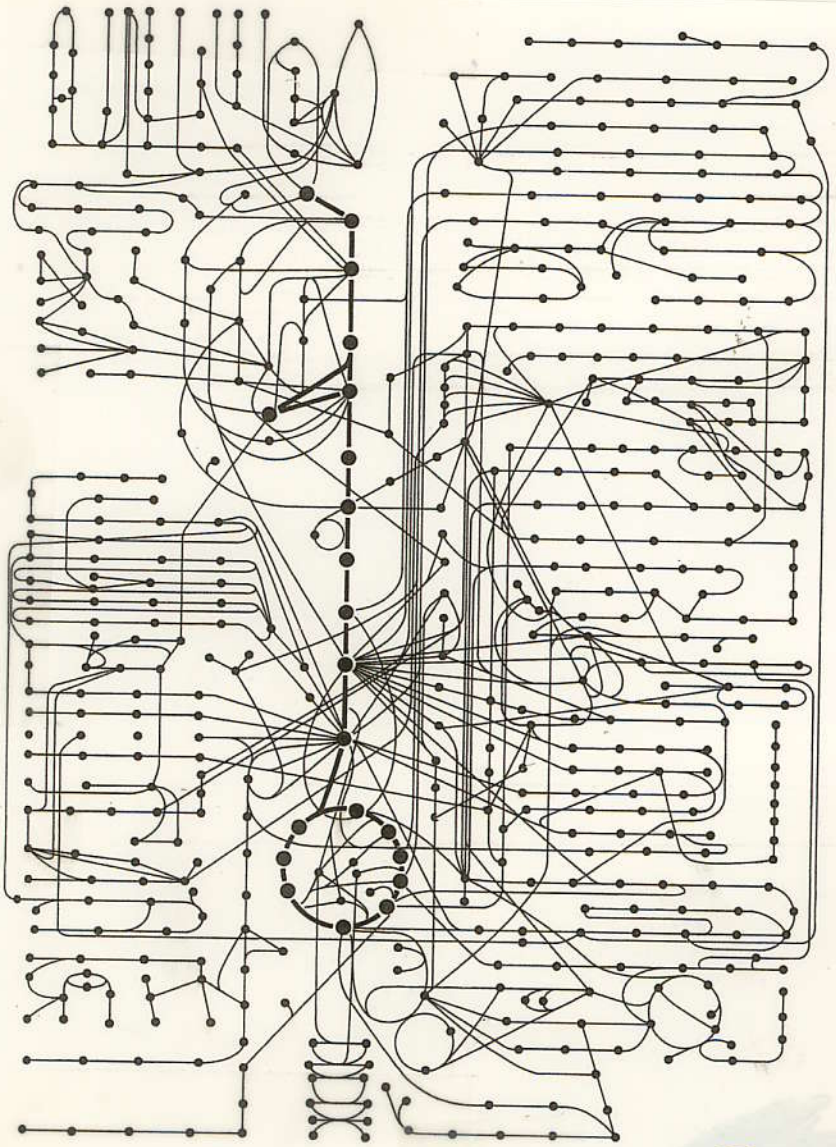


Figure 2–35 Some of the metabolic pathways and their interconnections in a typical cell. About 500 common metabolic reactions are shown diagrammatically, with each molecule in a metabolic pathway represented by a filled circle, as in the *yellow* box in Figure 2–34.

Goal: to find genes that encode enzymes that efficiently metabolize drugs
∴
no side effects

MOLECULAR SNP DETECTION

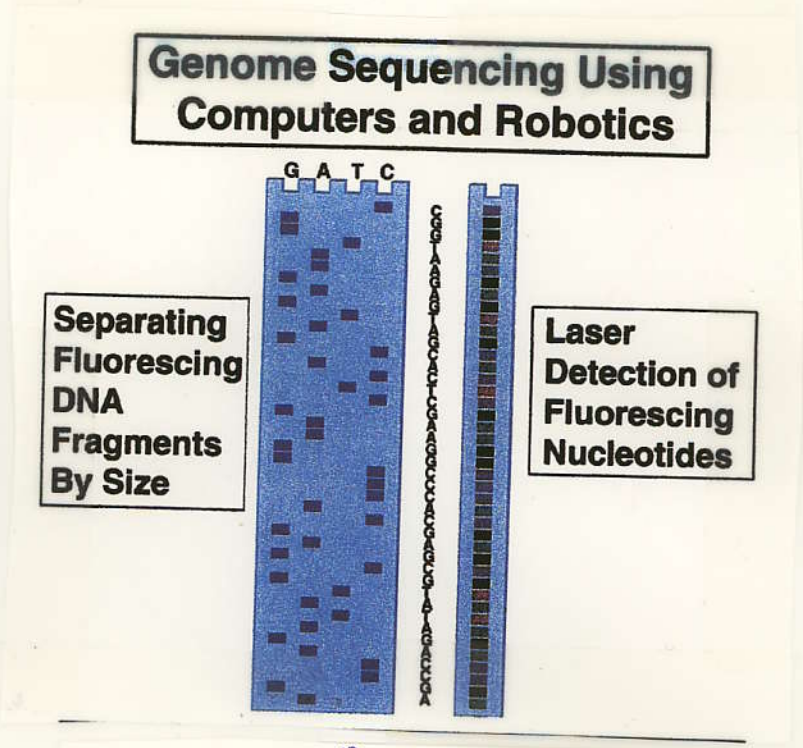
1. RFLPs → BLOTS
   → PCR
2. ASOs → PCR + Hybridization / "Blot"
3. Comparative Sequencing / Genes / Genome
4. Chips / Whole Genome / Individuals + Groups

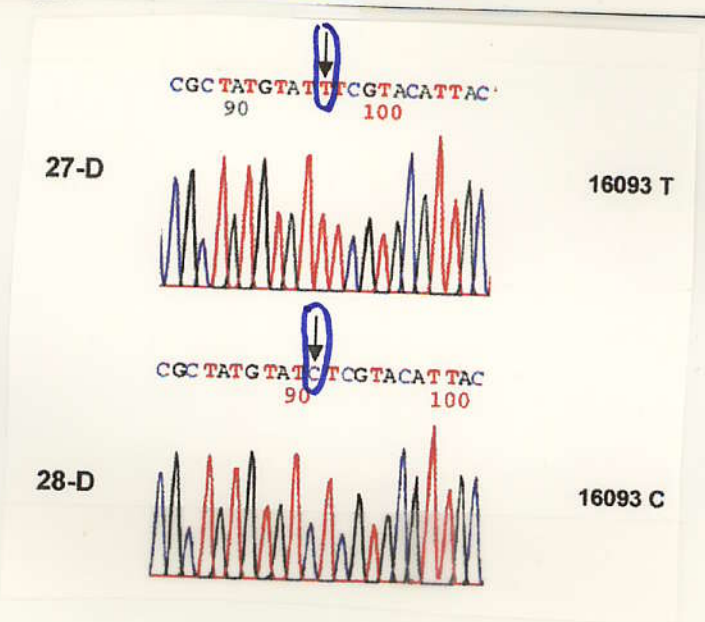
after SNPs identified by sequencing — + synthesized as ASOs on Chip!

DIRECT DNA SEQUENCING
&
CHIPS TO DETECT
SNPs

Gene & whole Genome Approaches

Detecting SNPs By Sequencing





Individual 1

Individual 2

Using Chips to Detect SNPs

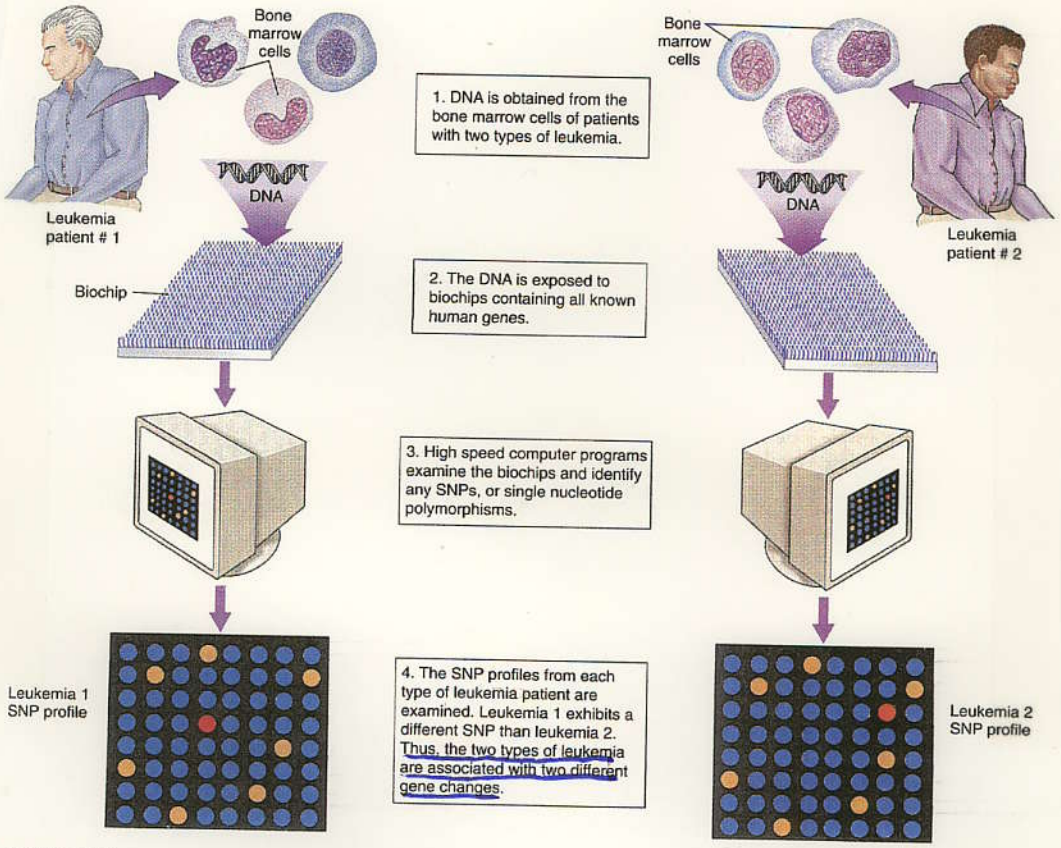


FIGURE 19.16
Biochips can help in identifying precise forms of cancer.

HAPLOTYPES OR SNPs on a Chromosome Inherited as a Unit

There are Millions of SNPs That Differ Among Individuals ....

.... But a small few reflect our Ancestry & "travel" in groups on chromosomes —

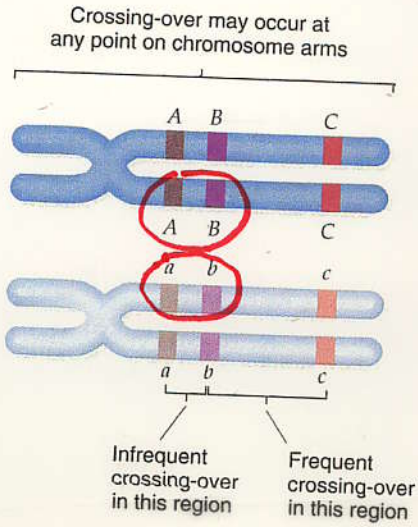
are linked & may show specific gene linkages!

Haplotypes!

CLOSELY-LINKED SNPs ARE INHERITED AS A UNIT

Figure 5.3

The relationship between recombination and map distance. The farther apart two genes are, the greater the number of possible sites for recombination. Thus, the probability of recombination occurring between genes *A* and *B* is much less than that between genes *B* and *C*. The percentage of recombinants can provide information about the relative genetic distance between two linked genes.

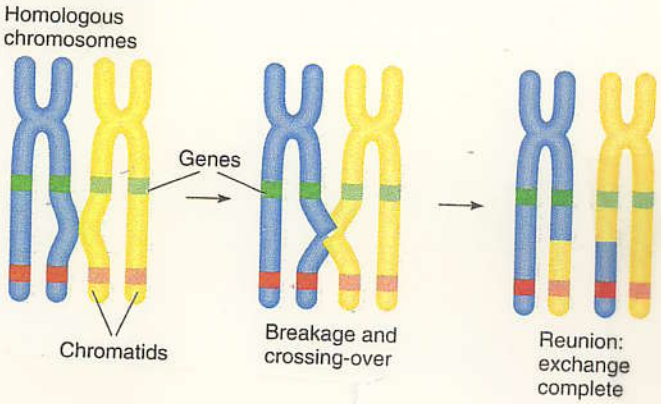


Haplotype
AB vs. ab

NO CROSSING OVER < 5kb
∴ HAPLOTYPE OF COMPLEX POLYMORPHIC LOCUS

Figure 5.2

Mechanism of crossing-over. A highly simplified diagram of a crossover between two nonsister chromatids during meiotic prophase, giving rise to recombinant (nonparental) combinations of linked genes.

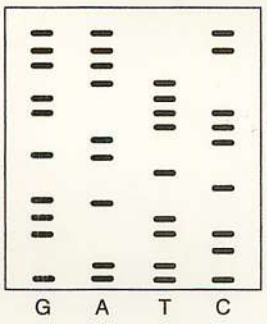


A Haplotype is a closely linked set of specific SNPs

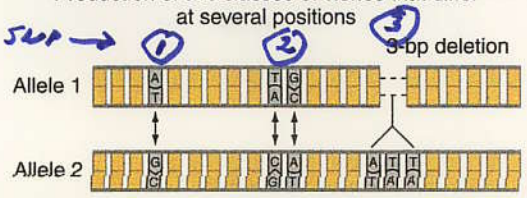
PCR amplification of HLAA locus from one person who is heterozygous for two complex haplotypes

Clone from PCR products

Sequence several clones to obtain at least one sequence from each of the two alleles.



Production of two classes of clones that differ at several positions



Mother

Father

3 SNP differences in each Chromosome

Figure 9.15 The variations associated with a complex haplotype are best defined by sequencing. Using automated protocols to sequence an entire polymorphic region is often the most rapid and accurate way to detect changes associated with polymorphic alleles at a complex locus.

They are ALWAYS inherited Together & Reflect Ancestry!

Haplotype
1 2 –
1' 2' 3

Mark Chromosome History – inherited As a Unit

HAPLOTYPES

TRACE ORIGINS OF HUMANS

IMMORTAL GENETIC VARIATION!

Complex Haplotypes

A contraction of the phrase "haploid genotype," the term **haplotype** refers to a specific combination of linked alleles in a cluster of related genes. Immunogeneticists often use it to describe the combination of alleles of the *major histocompatibility complex (MHC):* a large cluster of genes on human chromosome 6 that play a role in the immune response. With the resolving power to look at DNA at the level of nucleotides, "haplotype" now refers to any set of linked DNA changes along a chromosome. These changes could be in one or several genes, or in noncoding stretches. The **complex** refers to the multiple types of variation that can exist at alternative alleles, including more than one nucleotide substitution, a substitution in combination with a small deletion, duplication, or other insertion. Thus, a **complex haplotype** is a set of linked DNA variations along a chromosome, with the possibility of many differences between alternative alleles.

H1
H2
H1
H2
H1

HAPLOTYPE PATTERNS	
Person A	ATTGATCGGAT...CCATCGGA...CTAA
Person B	ATTGATAGGAT...CCAGCGGA...CTCA
Person C	ATTGATCGGAT...CCATCGGA...CTAA
Person D	ATTGATAGGAT...CCAGCGGA...CTCA
Person E	ATTGATCGGAT...CCATCGGA...CTAA

Building blocks. Persons B and D share a haplotype unlike the other three, characterized by three different SNPs.

CLOSELY LINKED SNPS

The International HapMap Project

The International HapMap Consortium*

*Lists of participants and affiliations appear at the end of the paper

The goal of the International HapMap Project is to determine the common patterns of DNA sequence variation in the human genome and to make this information freely available in the public domain. An international consortium is developing a map of these patterns across the genome by determining the genotypes of one million or more sequence variants, their frequencies and the degree of association between them, in DNA samples from populations with ancestry from parts of Africa, Asia and Europe. The HapMap will allow the discovery of sequence variants that affect common disease, will facilitate development of diagnostic tools, and will enhance our ability to choose targets for therapeutic intervention.

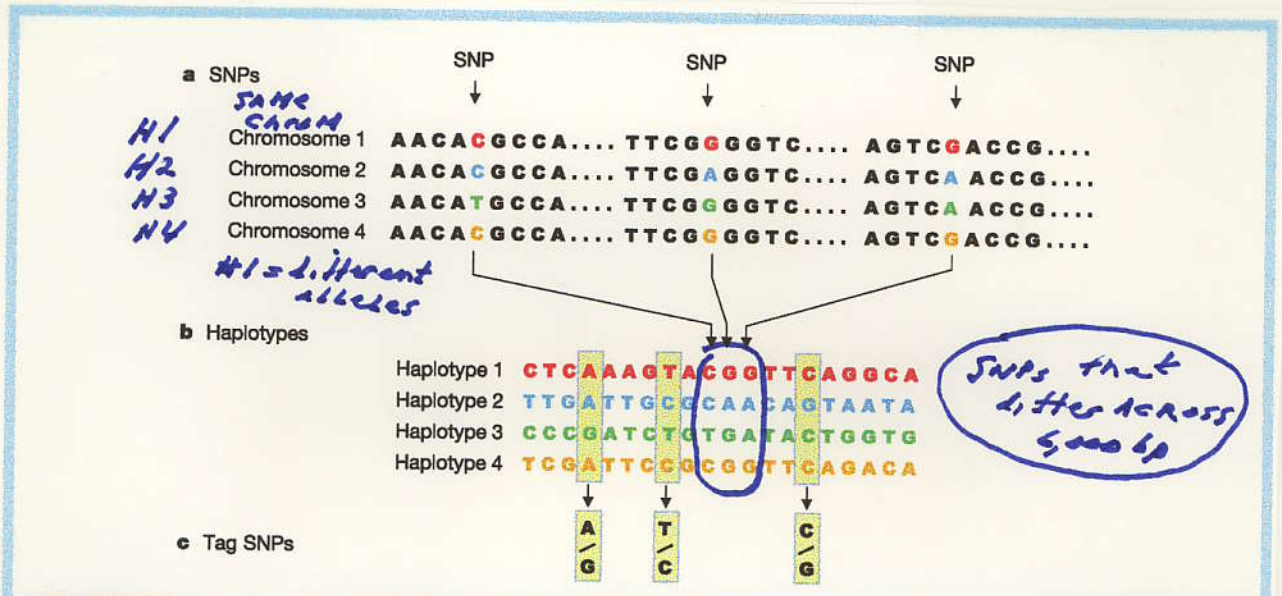


Figure 1 SNPs, haplotypes and tag SNPs. **a**, SNPs. Shown is a short stretch of DNA from four versions of the same chromosome region in different people. Most of the DNA sequence is identical in these chromosomes, but three bases are shown where variation occurs. Each SNP has two possible alleles; the first SNP in panel **a** has the alleles C and T. **b**, Haplotypes. A haplotype is made up of a particular combination of alleles at nearby SNPs. Shown here are the observed genotypes for 20 SNPs that extend across 6,000 bases of DNA. Only the variable bases are shown, including the three SNPs that are shown in panel **a**. For this region, most of the chromosomes in a population survey turn out to have haplotypes 1–4. **c**, Tag SNPs. Genotyping just the three tag SNPs out of the 20 SNPs is sufficient to identify these four haplotypes uniquely. For instance, if a particular chromosome has the pattern A–T–C at these three tag SNPs, this pattern matches the pattern determined for haplotype 1. Note that many chromosomes carry the common haplotypes in the population.

THE 0.1% THAT's different!

CORRELATE WITH SEQUENCE VARIANTS AFFECTING DISEASE

International HapMap Groups

FAR LEFT, L. GEORGIA/CORBIS; LEFT, CORBIS

R: RIGHT, J. SLATTER/CORBIS; FAR RIGHT, O. FRANKEN/CORBIS

World view: the HapMap initiative will gather genetic data from African, Asian and ancestrally European populations.

1. Northern European
2. Western European
3. Yoruba/Nigerian/African
4. Japanese + Han Chinese/Asian

"Group" Genetic Diversity to Disease & other Aspects of Biology

International HapMap Project

中文 | English | Français | 日本語 | Yoruba

About the HapMap

What is the HapMap?
Origins of Haplotypes
Health Benefits
Populations Sampled
Ethical Issues
Consent Forms
Community Advisory Groups(CAG)
Data Release Policy
Guidelines For Data Use
Guidelines For Referring to HapMap Populations

Project Information

About the Project
HapMap Publications
HapMap Tutorial
HapMap Mailing List
HapMap Project Participants
HapMap Mirror Site in Japan

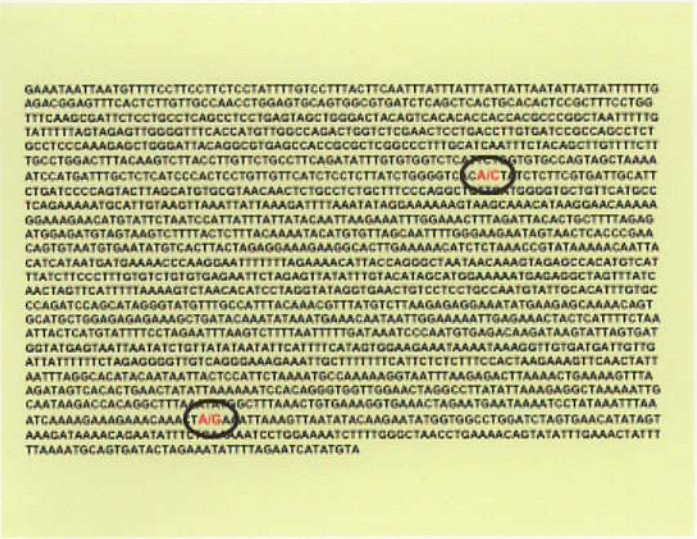
Useful Links

HapMap Project Press Release
NHGRI HapMap Page
NCBI Variation Database (dbSNP)
Japanese SNP Database (JSNP)

What Is the HapMap?

The HapMap is a catalog of common genetic variants that occur in human beings. It describes what these variants are, where they occur in our DNA, and how they are distributed among people within populations and among populations in different parts of the world. The International HapMap Project is not using the information in the HapMap to establish connections between particular genetic variants and diseases. Rather, the Project is designed to provide information that other researchers can use to link genetic variants to the risk for specific illnesses, which will lead to new methods of preventing, diagnosing, and treating disease.

The DNA in our cells



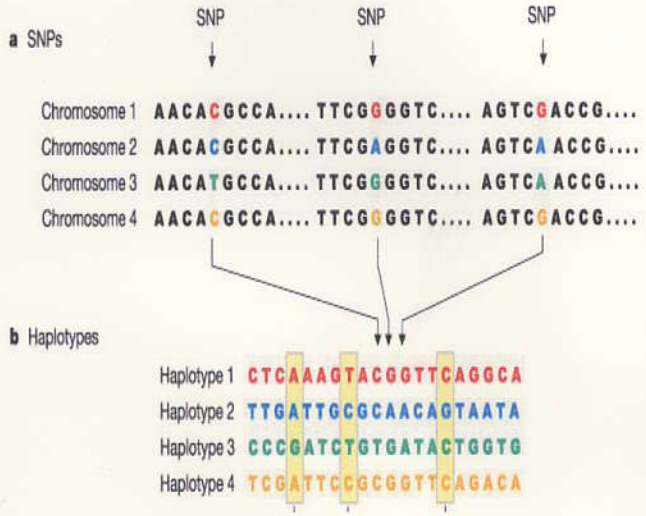


Figure 1: When DNA sequences on a part of chromosome 7 from two random individuals are compared, two single nucleotide polymorphisms (SNPs) occur in about 2,200 nucleotides.

contains long chains of four chemical building blocks -- adenine, thymine, cytosine, and guanine, abbreviated A, T, C, and G. More than 6 billion of these chemical bases, strung together in 23 pairs of chromosomes, exist in a human cell. (See **http://www.dnaftb.org/dnaftb/** for basic information about genetics.) These genetic sequences contain information that influences our physical traits, our likelihood of suffering from disease, and the responses of our bodies to substances that we encounter in the environment.

Figure 2: The construction of the HapMap occurs in three steps. (a) Single nucleotide polymorphisms(SNPs) are identified in DNA samples from multiple indivduals. (b)Adjacent SNPs that are inherited together are compiled into "haplotypes." (c)"Tag" SNPs within haplotypes are identified that uniquely identify those haplotypes. By genotyping the three tag SNPs shown in this figure, researchers can identify which of the four haplotypes shown here are present in each individual.

The genetic sequences of different people are remarkably similar. When the chromosomes of two humans are compared, their DNA sequences can be identical for hundreds of bases. But at about one in every 1,200 bases, on average, the sequences will differ (Figure 1). One person might have an A at that location, while another person has a G, or a person might have extra bases at a given location or a missing segment of DNA. Each distinct "spelling" of a chromosomal region is called an allele, and a collection of alleles in a person's chromosomes is known as a genotype.

Differences in individual bases are by far the most common type of genetic variation. These genetic differences are known as single nucleotide polymorphisms, or SNPs (pronounced "snips"). By identifying most of the approximately 10 million SNPs estimated to occur commonly in the human genome, the International HapMap Project is identifying the basis for a large fraction of the genetic diversity in the human species.

For geneticists, SNPs act as markers to locate genes in DNA sequences. Say that a spelling change in a gene increases the risk of suffering from high blood pressure, but researchers do not know where in our chromosomes that gene is located. They could compare the SNPs in people who have high blood pressure with the SNPs of people who do not. If a particular SNP is more common among people with hypertension, that SNP could be used as a pointer to locate and identify the gene involved in the disease.

However, testing all of the 10 million common SNPs in a person's chromosomes would be extremely expensive. The development of the HapMap will enable geneticists to take advantage of how SNPs and other genetic variants are organized on chromosomes. Genetic variants that are near each other tend to be inherited together. For example, all of the people who have an A rather than a G at a particular location in a chromosome can have identical genetic variants at other SNPs in the chromosomal region surrounding the A. These regions of linked variants are known as haplotypes (Figure 2).

In many parts of our chromosomes, just a handful of haplotypes are found in humans. [**See The Origins of Haplotypes.**] In a given population, 55 percent of people may have one version of a haplotype, 30 percent may have another, 8 percent may have a third, and the rest may have a variety of less common haplotypes. The International HapMap Project is identifying these common haplotypes in four populations from different parts of the world. It also is identifying "tag" SNPs that uniquely identify these haplotypes. By testing an individual's tag SNPs (a process known as genotyping), researchers will be able to identify the collection of haplotypes in a person's DNA. The number of tag SNPs that contain most of the information about the patterns of genetic variation is estimated to be about 300,000 to 600,000, which is far fewer than the 10 million common SNPs.

Once the information on tag SNPs from the HapMap is available, researchers will be able to use them to locate genes involved in medically important traits. Consider the researcher trying to find genetic variants associated with high blood

pressure. Instead of determining the identity of all SNPs in a person's DNA, the researcher would genotype a much smaller number of tag SNPs to determine the collection of haplotypes present in each subject. The researcher could focus on specific candidate genes that may be associated with a disease, or even look across the entire genome to find chromosomal regions that may be associated with a disease. If people with high blood pressure tend to share a particular haplotype, variants contributing to the disease might be somewhere within or near that haplotype.

Please send questions and comments on website to help@hapmap.org

International HapMap Project

Which Populations Are Being Sampled

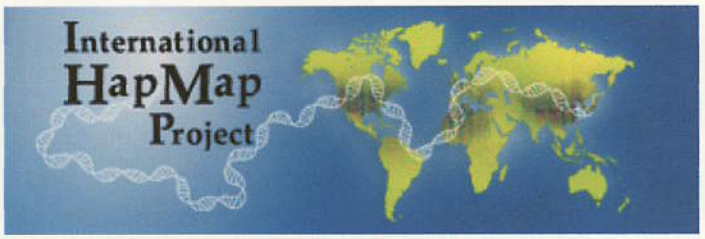
The International HapMap Project is analyzing DNA from populations with African, Asian, and European ancestry. Together, these DNA samples should enable HapMap researchers to identify most of the common haplotypes that exist in populations worldwide. [**See What Is the HapMap?**]

Because of the history of the human species, most of the common haplotypes in human chromosomes occur in all human populations. [**See The Origin of Haplotypes.**] However, any given haplotype may be more common in one population and less common in another, and newer haplotypes may be found in just a single population. Efficiently choosing the tag SNPs needed to identify haplotypes therefore requires looking at haplotype frequencies in multiple populations. Also, genetic data from more than one population will enhance the ability of researchers to study the genetic contributions to diseases that are more or less prevalent in different groups.

The DNA samples for the HapMap have come from a total of 270 people. The Yoruba people of Ibadan, Nigeria, provided 30 sets of samples from two parents and an adult child (each such set is called a trio). In Japan, 45 unrelated individuals from the Tokyo area provided samples. In China, 45 unrelated individuals from Beijing provided samples. Thirty U.S. trios provided samples, which were collected in 1980 from U.S. residents with northern and western European ancestry by the Centre d'Etude du Polymorphisme Humain (CEPH).

The blood samples are being converted into cell lines, which are used to make DNA, by the non-profit Coriell Institute for Medical Research. [See **http://locus.umdnj.edu/nigms** for more information.] Coriell provides DNA and cell lines from the samples for research projects that have been approved by the appropriate ethics committees. The samples and cell lines are not linked to any individual in the populations studied. However, the samples and cell lines are identified as coming from one of the four populations participating in the study, which raises ethical issues associated with conducting genetic research in named populations. [**See How Are Ethical Issues Being Addressed?** and **Guidelines for Referring to the HapMap Populations in Publications and Presentations**.]

To assess how much additional information would be gained by genotyping other populations, haplotypes in a set of chromosomal regions are being analyzed in samples from several additional populations.



International HapMap Project

How Will the HapMap Benefit Human Health?

The International HapMap Project will benefit human health by providing an extensive resource that researchers can use to discover the genetic variants involved in disease and individual responses to therapeutic agents. Once such variants have been discovered, researchers can learn much more about the origins of illnesses and about ways to prevent, diagnose, and treat those illnesses.

The goal of the Project is not to identify these disease-related genes directly. Rather, by identifying haplotypes, the HapMap provides a tool that can be used in what are called association studies. For these studies, researchers will compare the haplotypes in individuals with a disease to the haplotypes of a comparable group of individuals without a disease (the controls). If a particular haplotype occurs more frequently in affected individuals compared with controls, a gene influencing the disease may be located within or near that haplotype.

Common diseases such as cancer, stroke, heart disease, diabetes, depression, and asthma usually result from the combined effects of a number of genetic variants and environmental factors. According to an idea known as the common disease-common variant hypothesis, the risk of contracting common diseases is influenced by genetic variants that are relatively common in populations. Not enough data are yet available to evaluate the generality of this hypothesis, but more and more widely distributed genetic variants associated with common diseases are being discovered, including variants that contribute to autoimmune diseases, schizophrenia, diabetes, asthma, stroke, and heart attacks. One of the many benefits of the International HapMap Project will be the use of the HapMap to learn more about the links between these common disorders and our genes.

Knowledge derived from use of the HapMap also will result in advances that are difficult to predict today. Medical treatments could be customized, based on a patient's genetic make-up, to maximize effectiveness and minimize side effects. Genetic variants contributing to longevity or resistance to disease could be identified, leading to new therapies with widespread benefits. As with any new body of knowledge, the HapMap is likely to lead to both new challenges and to unexpected and unprecedented opportunities.

Human pigmentation genetics: the difference is only skin deep

Richard A. Sturm,[1]* Neil F. Box,[1] and Michele Ramsay[2]

Summary

There is no doubt that visual impressions of body form and color are important in the interactions within and between human communities. Remarkably, it is the levels of just one chemically inert and stable visual pigment known as melanin that is responsible for producing all shades of humankind. Major human genes involved in its formation have been identified largely using a comparative genomics approach and through the molecular analysis of the pigmentary process that occurs within the melanocyte. Three classes of genes have been examined for their contribution to normal human color variation through the production of hypopigmented phenotypes or by genetic association with skin type and hair color. The MSH cell surface receptor and the melanosomal P-protein are the two most obvious candidate genes influencing variation in pigmentation phenotype, and may do so by regulating the levels and activities of the melanogenic enzymes tyrosinase, TRP-1 and TRP-2. *BioEssays* **20**:712–721, 1998.

TABLE 1. Human Pigmentation Genes

Gene symbol	Mouse homologue	Chromosome	Phenotype	Protein	Function/ activity
TYR	Albino (c)	11q14–21	OCA1	Tyrosinase	Tyrosine hydroxylation; DOPA oxidase
TYRP1	Brown (b)	9p23	OCA3/ROCA	TRP-1	DHICA oxidase
TYRP2	Slaty (slt)	13q31–32	Unknown	TRP-2	Dopachrome tautomerase
P	Pink-eyed dilute (p)	15q11.2–12	OCA2, BOCA	P-protein	Melanosomal transmembrane protein
MC1R	Extension (e)	16q24.3	Red hair	MSHR	G-protein-coupled receptor

(+)
SLC24A5

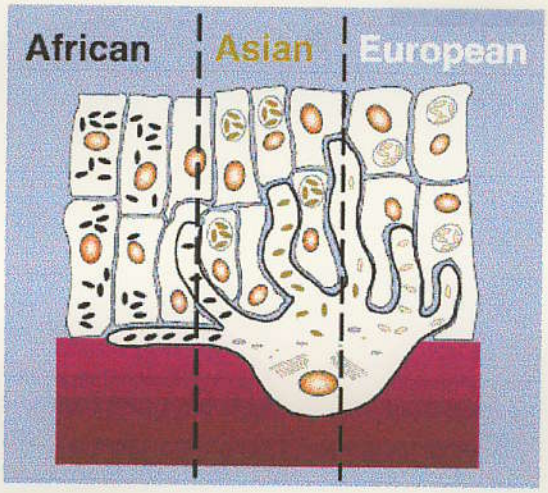


Figure 1. Variation in melanosome structure and distribution in different groups. A single skin melanocyte cell interdigitating with keratinocyte cells is partitioned into three sections. Shown within the melanocyte are the four stages of melanosome formation from budding from the Golgi apparatus, to the fully pigmented stage IV melanosomes migrating up the dendritic processes of the cell and secreted into the keratinocytes. In African populations, the melanosomes remain as singular heavily pigmented particles while in Asians and Europeans the melanosomes cluster in membrane bound organelles giving different skin complexions.

SLC24A5, a Putative Cation Exchanger, Affects Pigmentation in Zebrafish and Humans

Rebecca L. Lamason,[1*] Manzoor-Ali P.K. Mohideen,[1†] Jason R. Mest,[1] Andrew C. Wong,[1‡] Heather L. Norton,[6] Michele C. Aros,[1] Michael J. Jurynec,[8] Xianyun Mao,[6] Vanessa R. Humphreville,[1§] Jasper E. Humbert,[2,9] Soniya Sinha,[2] Jessica L. Moore,[1‖] Pudur Jagadeeswaran,[10] Wei Zhao,[3] Gang Ning,[7] Izabela Makalowska,[7] Paul M. McKeigue,[11] David O'Donnell,[11] Rick Kittles,[12] Esteban J. Parra,[13] Nancy J. Mangini,[14] David J. Grunwald,[8] Mark D. Shriver,[6] Victor A. Canfield,[4] Keith C. Cheng[1,4,5¶]

Lighter variations of pigmentation in humans are associated with diminished number, size, and density of melanosomes, the pigmented organelles of melanocytes. Here we show that zebrafish *golden* mutants share these melanosomal changes and that *golden* encodes a putative cation exchanger slc24a5 (nckx5) that localizes to an intracellular membrane, likely the melanosome or its precursor. The human ortholog is highly similar in sequence and functional in zebrafish. The evolutionarily conserved ancestral allele of a human coding polymorphism predominates in African and East Asian populations. In contrast, the variant allele is nearly fixed in European populations, is associated with a substantial reduction in regional heterozygosity, and correlates with lighter skin pigmentation in admixed populations, suggesting a key role for the *SLC24A5* gene in human pigmentation.

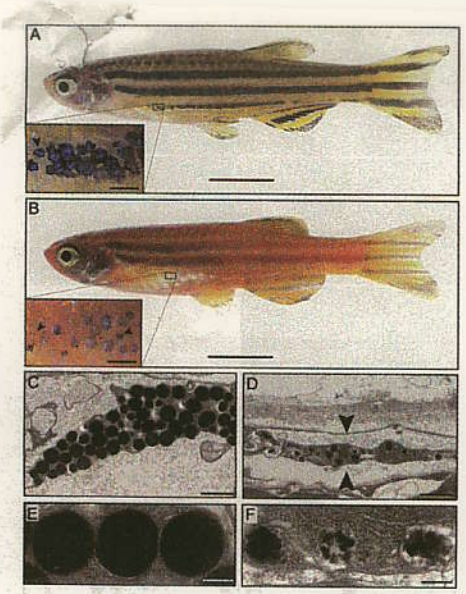


Fig. 1. Phenotype of *golden* zebrafish. Lateral views of adult wild-type (A) and *golden* (B) zebrafish. Insets show melanophores (arrowheads). Scale bars, 5 mm (inset, 0.5 mm). gol^{b1} mutants have melanophores that are, on average, smaller, more pale, and transparent. Transmission electron micrographs of skin melanophore from 55-hpf wild-type (C and E) and gol^{b1} (D and F) larvae. gol^{b1} skin melanophores (arrowheads show edges) are thinner and contain fewer melanosomes than do those of wild type. Melanosomes of gol^{b1} larvae are fewer in number, smaller, less-pigmented, and irregular compared with wild type. Scale bars in (C) and (D), 1000 nm; in (E) and (F), 200 nm.

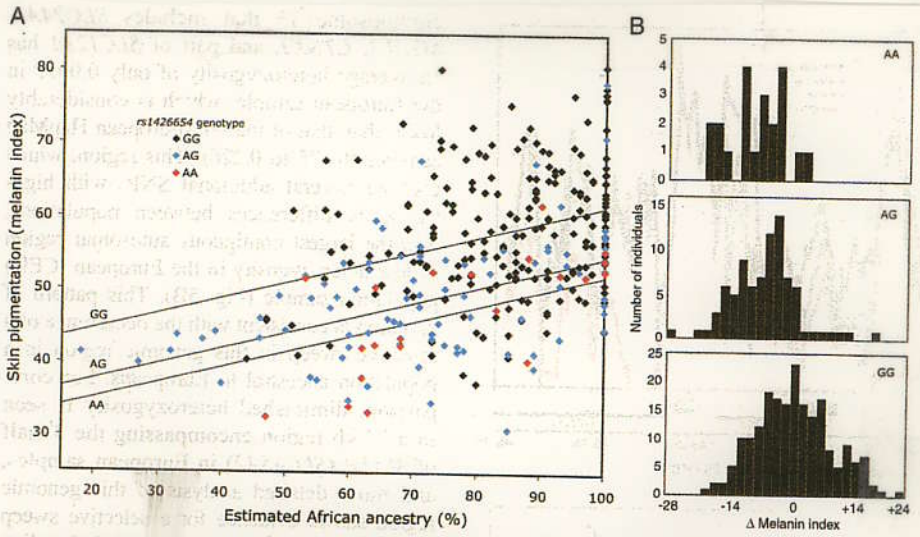
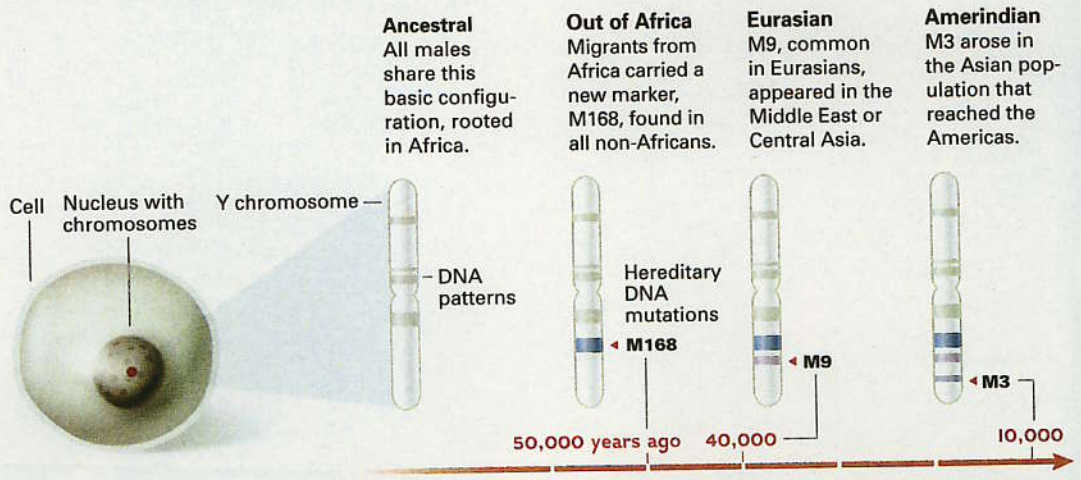


Fig. 6. Effect of *SLC24A5* genotype on pigmentation in admixed populations. (A) Variation of measured pigmentation with estimated ancestry and *SLC24A5* genotype. Each point represents a single individual; *SLC24A5* genotypes are indicated by color. Lines show regressions, constrained to have equal slopes, for each of the three genotypes. (B) Histograms showing the distribution of pigmentation after adjustment for ancestry for each genotype. Values shown are the difference between the measured melanin index and the calculated GG regression line ($y = 0.2113x + 30.91$). The corresponding uncorrected histograms are shown in fig. S7. Mean and SD (in parentheses) are given as follows: for GG, 0 (8.5), $n = 202$ individuals; for AG, –7.0 (7.4), $n = 85$; for AA, –9.6 (6.4), $n = 21$.

HAPLOTYPES CAN BE USED TO TRACE ORIGINS OF HUMAN POPULATIONS

History on a Chromosome

Genetic mutations act as markers, tracing a journey through time. The earliest known mutation to spread outside Africa is M168, which arose some 50,000 years ago. This graphic shows the Y chromosome of a Native American man with various mutations including M168, proving his African ancestry.



68 NATIONAL GEOGRAPHIC • MARCH 2006

NGM ART

Note - looking at novel alleles in populations! And their frequencies!

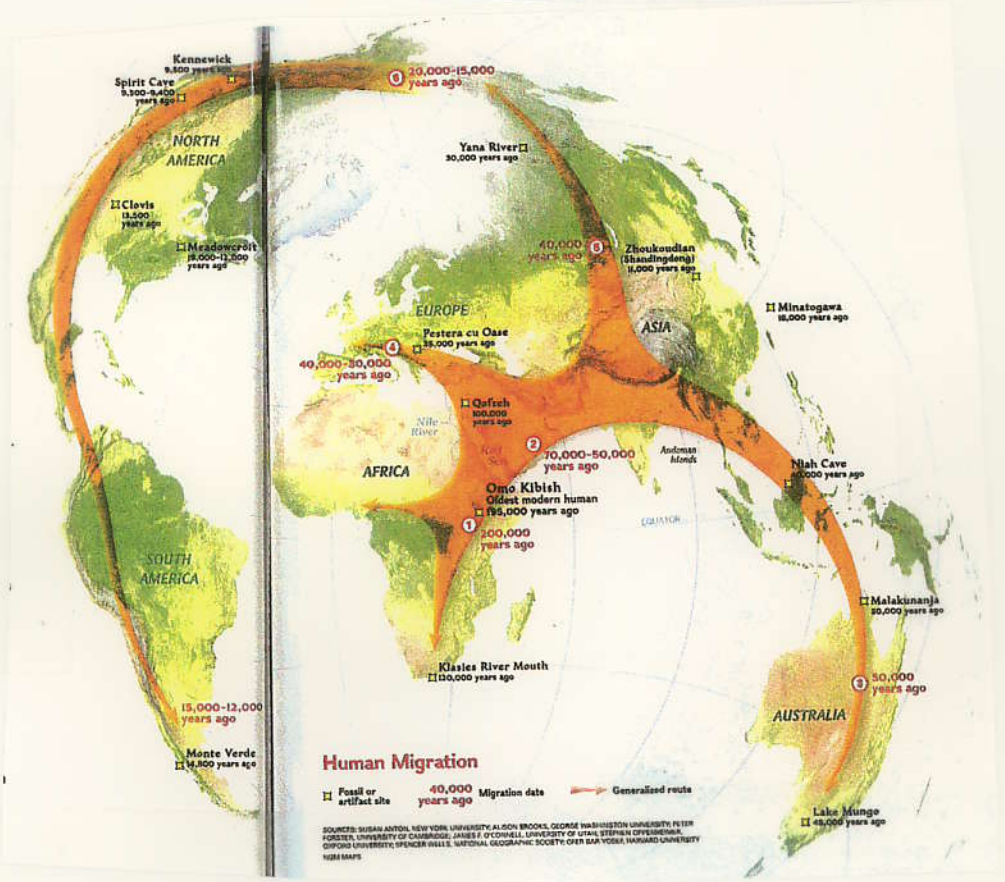
e.g. A vs. a

B vs b

C vs c

Markers MUST have different "forms"!

What Genes and Fossils Tell Us



1 African Cradle

Most paleoanthropologists and geneticists agree that modern humans arose some 200,000 years ago in Africa. The earliest modern human fossils were found in Omo Kibish, Ethiopia. Sites in Israel hold the earliest evidence of modern humans outside Africa, but that group went no farther, dying out about 90,000 years ago.

2 Out of Africa

Genetic data show that a small group of modern humans left Africa for good 70,000 to 50,000 years ago and eventually replaced all earlier types of humans, such as Neandertals. All non-Africans are the descendants of these travelers, who may have migrated around the top of the Red Sea or across its narrow southern opening.

3 The First Australians

Discoveries at two ancient sites—artifacts from Malakunanja and fossils from Lake Mungo—indicate that modern humans followed a coastal route along southern Asia and reached Australia nearly 50,000 years ago. Their descendants, Australian Aborigines, remained genetically isolated on that island continent until recently.

4 Early Europeans

Paleoanthropologists long thought that the peopling of Europe followed a route from North Africa through the Levant. But genetic data show that the DNA of today's western Eurasians resembles that of people in India. It's possible that an inland migration from Asia seeded Europe between 40,000 and 30,000 years ago.

5 Populating Asia

Around 40,000 years ago, humans pushed into Central Asia and arrived on the grassy steppes north of the Himalaya. At the same time, they traveled through Southeast Asia and China, eventually reaching Japan and Siberia. Genetic clues indicate that humans in northern Asia eventually migrated to the Americas.

6 Into the New World

Exactly when the first people arrived in the Americas is still hotly debated. Genetic evidence suggests it was between 20,000 and 15,000 years ago, when sea levels were low and land connected Siberia to Alaska. Ice sheets would have covered the interior of North America, forcing the new arrivals to travel down the west coast.

WATCH "BLACKBEARD" ON NG CHANNEL, SUNDAY, MARCH 12, 8 P.M. ET/9 P.M. PT

NATIONALGEOGRAPHIC.COM/MAGAZINE MARCH 2006

NATIONAL GEOGRAPHIC

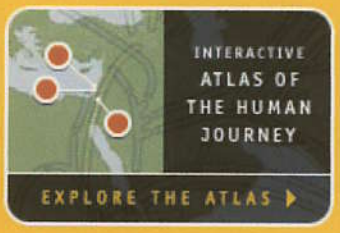
The Greatest Journey Ever Told

THE TRAIL OF OUR DNA

54

A LANDMARK STUDY OF THE HUMAN JOURNEY

Who was **your** first ancestor? New DNA studies say that all humans descended from an African ancestor who lived only 60,000 years ago. Uncover the specific paths that led from him to you—the ultimate human history, as written in our genes.



NEWS AND RESOURCES

- **Millions of Men May Be Descended From Irish King, Study Says**
- **More Related News**
- **Related Web Resources**

GENETICS OVERVIEW

Delve inside and explore the basics of genetics, from chromosomes to natural selection.

YOUR GENETIC JOURNEY

Explore your own genetic journey with Dr. Spencer Wells. DNA analysis includes a depiction of your ancient ancestors and an interactive map tracing your genetic lineage around the world and through the ages.

Enter Your Genographic Project Kit ID Here

LOG IN

Remember my log in (optional)

Question of the Week

- **What happens once I order a Genographic Project Participation Kit?**

ALSO SEE

- **Buy the Participation Kit**
- **Help Support the Genographic Project Field Research**

THE WAITT FAMILY FOUNDATION

Global field research supported by the Waitt Family Foundation

NATIONAL GEOGRAPHIC

A research partnership of National Geographic and IBM

The Genographic Project

Published 2 weeks, 1 day ago

Two months ago my father told me that he had just sent a cheek swab sample to the National Geographic Genographic Project, which, at the time, I had never heard of. To have the website tell it, the goal of the project is to "understand the human journey — where we came from and how we got to where we live today." The data collected "will map world migratory patterns dating back some 150,000 years and will fill in the huge gaps in our knowledge of humankind's migratory history."

The participation kit is only $100 and it's money well spent if you ask me.

After having read both Jared Diamond's Guns, Germs, and Steel: The Fates of Human Societies and Steve Olson's Mapping Human History: Genes, Race, and Our Common Origins, I couldn't wait for my father to receive the results — I'm completely fascinated by this stuff.

The results

> To be clear—these tests are not conventional genealogy. Your results will not provide names for your personal family tree or tell you where your great grandparents lived. Rather, they will indicate the maternal or paternal genetic markers your deep ancestors passed on to you and the story that goes with those markers.

A few days ago he got the results and shared the online account with me. They're utterly fascinating. The picture below shows my "ancestral journey." At the website, this picture is interactive — clicking on a marker will tell you the story behind it.

In addition to this migratory map, you're also given your genetic sequence (not reproduced here) and told how to interpret it in light of these mass migratory patterns. Finally, you're told your "genetic history" (reproduced below), which basically summarizes where and when your haplogroup originated, how they got

HOW TO PARTICIPATE

Public participation, including yours, is critical to the Genographic Project's success. Here's how you can get involved:

Purchasing a Public Participation Kit will fund important research around the world—and open the door to the ancient past of your own genetic background.

With a simple and painless cheek swab you can sample your own DNA. You'll submit the sample through our secure, private, and completely anonymous system, then log on to the project Web site to track your personal results online.

This is not a genealogy test and you won't learn about your great grandparents. You will learn, however, of your deep ancestry, the ancient genetic journeys and physical travels of your distant relatives.

To insure total anonymity you will be identified at all times only by your kit number, not by your name. There is no record, no database that links test results with the names of their contributors. If you lose the kit number there will be no way to access your genetic results.

The kit includes the following elements:

- **Buccal swab kit**
- **Multimedia DVD**
- **Exclusive National Geographic Genographic Map**
- **"Quick Start" card**
- **Genographic Project Brochure**
- **Self-addressed envelope**
- **Confidential Genographic Project ID Number (GPID)**

The purchase price also includes the cost of the testing and analysis. International participants please see note below.

As your own genetic ancestry is revealed you'll also see worldwide samples map humankind's shared genetic background around the world and through the ages.

If you'd like to contribute your own results to the project's global database you'll be asked to answer a dozen "phenotyping" questions that will help place your DNA in cultural context.

This process is optional and completely anonymous, but it's also important. Each of us has a part

USES OF THE HAPMAP

Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes

M. F. Hammer*†‡, A. J. Redd*†, E. T. Wood*†, M. R. Bonner*, H. Jarjanazi*, T. Karafet*, S. Santachiara-Benerecetti¶, A. Oppenheim‖, M. A. Jobling**, T. Jenkins††, H. Ostrer‡‡, and B. Bonné-Tamir§

*Laboratory of Molecular Systematics and Evolution, University of Arizona, Tucson, AZ 85721; ¶Department of Genetics, Università degli Studi di Pavia, Pavia 27100, Italy; ‖Hadassah Medical School, Hebrew University of Jerusalem, Jerusalem 91120, Israel; **Department of Genetics, University of Leicester, Leicester LE1 7RH, England; ††SAMIR, University of Witwatersrand, Johannesburg 2000, South Africa; ‡‡Department of Pediatrics, New York University Medical Center, New York, NY 10016; and §Department of Human Genetics, Sackler School of Medicine, Ramat Aviv 69978, Israel

Communicated by Arno G. Motulsky, University of Washington, Seattle, WA, March 15, 2000 (received for review November 17, 1999)

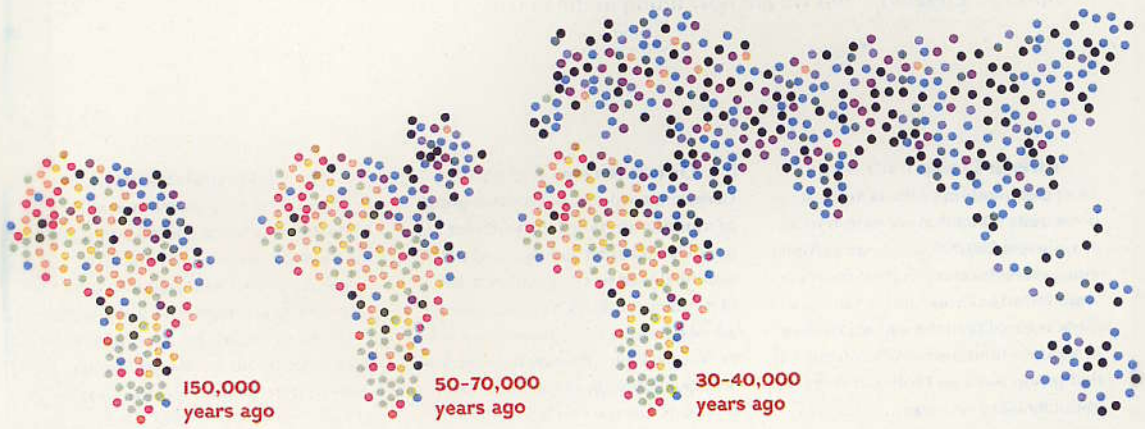
Haplotypes constructed from Y-chromosome markers were used to trace the paternal origins of the Jewish Diaspora. A set of 18 biallelic polymorphisms was genotyped in 1,371 males from 29 populations, including 7 Jewish (Ashkenazi, Roman, North African, Kurdish, Near Eastern, Yemenite, and Ethiopian) and 16 non-Jewish groups from similar geographic locations. The Jewish populations were characterized by a diverse set of 13 haplotypes that were also present in non-Jewish populations from Africa, Asia, and Europe. A series of analyses was performed to address whether modern Jewish Y-chromosome diversity derives mainly from a common Middle Eastern source population or from admixture with neighboring non-Jewish populations during and after the Diaspora. Despite their long-term residence in different countries and isolation from one another, most Jewish populations were not significantly different from one another at the genetic level. Admixture estimates suggested low levels of European Y-chromosome gene flow into Ashkenazi and Roman Jewish communities. A multidimensional scaling plot placed six of the seven Jewish populations in a relatively tight cluster that was interspersed with Middle Eastern non-Jewish populations, including Palestinians and Syrians. Pairwise differentiation tests further indicated that these Jewish and Middle Eastern non-Jewish populations were not statistically different. The results support the hypothesis that the paternal gene pools of Jewish communities from Europe, North Africa, and the Middle East descended from a common Middle Eastern ancestral population, and suggest that most Jewish communities have remained relatively isolated from neighboring non-Jewish communities during and after the Diaspora.

5P

MOST GENETIC DIVERSITY ORIGINATED in the founder populations to Modern HUMANS!

Diverse From the Start

The diversity of genetic markers is greatest in Africa (multicolored dots in map), indicating it was the earliest home of modern humans. Only a handful of people, carrying a few of the markers, walked out of Africa (center) and, over tens of thousands of years, seeded other lands (right). "The genetic makeup of the rest of the world is a subset of what's in Africa," says Yale geneticist Kenneth Kidd.



ARE There HUMAN
"Races?"

What is the History
& the Biology?

RACE IS PRIMARILY A SOCIOLOGICAL CONCEPT THAT HAS CAUSED MUCH HUMAN SUFFERING

Race is a largely non-biological concept confounded by misunderstanding and a long history of prejudice. The relationship of genomics to the concepts of race and ethnicity has to be considered within complex historical and social contexts.

Most variation in the genome is shared between all populations, but certain alleles are more frequent in some populations than in others, largely as a result of history and geography. Use of genetic data to define racial groups, or of racial categories to classify biological traits, is prone to misinterpretation. To minimize such misinterpretation, the biological and sociocultural factors that interrelate genetics with constructs of race and ethnicity need to be better understood and communicated within the next few years.

This will require research on how different individuals and cultures conceive of race, ethnicity, group identity and self-identity, and what role they believe genes or other biological factors have. It will also require a critical examination of how the scientific community understands and uses these concepts in designing research and presenting findings, and of how the media report these. Also necessary is widespread education about the biological meaning and limitations of research findings in this area (Box 6) and the formulation and adoption o public-policy options that protect agains genomics-based discrimination or maltreatment (see Grand Challenge III-1).

GENOME PROJECT GOAL

① MOST VARIATION SHARED

② DIFFERENCES DUE TO MIGRATIONS & GEOGRAPHIC ISOLATION

③ MOST VARIATION SHARED BY GROUPS!!

BASED ON A VERY FEW GENES THAT VARY BETWEEN GROUPS FAR MORE THAN MAJORITY – VAST MAJORITY OF OTHER GENES

A genetic melting-pot

Marcus W. Feldman, Richard C. Lewontin and Mary-Claire King

Race as a biological concept has had a variety of meanings. In the taxonomic literature, a race is any distinguishable type within a species, such as dark-bellied and light-bellied variants of small mammals. In 1937, Theodosius Dobzhansky introduced the idea of geographical races — populations of species that differ in the frequencies of one or more genetic variants. But as no two populations have identical gene frequencies at variable (polymorphic) loci, Dobzhansky's definition of race becomes synonymous with that of population.

The classical definition of race, as applied to our species, is based on phenotypes such as skin colour, facial features and hair form that clearly differ between native inhabitants of different regions of the world. An underlying assumption is that all of these defining features (all largely genetic traits, although few of their genes have been identified) are characteristic of the genome in general. In other words, just as there are large differences between races in genes for skin colour, so there should be large genetic differences between races in general. In the previous absence of data to confirm or deny this assumption, it was not an unreasonable one to make.

But recent studies of genetic diversity indicate that the genes underlying the phenotypic differences used to assign race categories are atypical, in that they vary between races much more than genes in general. Together, the iconic features of race correlate well with continent of origin but do not reflect genome-wide differences between groups.

Discussion has arisen over the implications of these findings for the utility of racial classification in medical practice. The issue of whether race is a biologically useful or even meaningful concept when applied to humans in a medical context is controversial — holders of opposing views each claim to have evidence to support them. But there is no contradiction between these two well-substantiated bodies of data, as they actually deal with two different questions that have become confused with one another.

STAPLETON COLLECTION/CORBIS

Human migration, as depicted in Charles Hunt's *Landing at Madras*, blurs the boundaries of race.

The first question is: "Is it possible to find DNA sequences that differ sufficiently between populations to allow correct assignment of major geographical origin with high probability?" The answer to this question is yes, as shown by studies of genetic polymorphisms and by universal personal experience.

The second question is: "What fraction of human genetic variation, whether based on protein-coding genes or other sequences, falls within geographically separated populations, and what fraction occurs between these populations?" The answer to this question is that most genetic diversity occurs within groups, and that very little is found between them.

Why this apparent paradox? The answer is that genes that are geographically distinctive in their frequencies are not typical of the human genome in general.

It has been suggested that racial categorization has a valid role in good medical practice because many medically important genes vary between populations from different regions. But although knowing a patient's ancestry is often extremely useful in diagnosis and treatment, race is both too broad and too narrow a definition of ancestry to be biologically useful.

For any species, definitions of race can lose their discriminating power when individuals migrate to different regions and mate with their counterparts there. Among humans, large-scale migrations between continents — particularly through European colonial expansion and the commercial slave trade — has resulted in matings of individuals from different continents and the creation of new populations, especially in the Western Hemisphere and Oceania. Many people thus have ancestry from more than one major geographical region, meaning that the association of phenotype and geography breaks down.

For example, sickle-cell disease, which is often thought to be an African trait, is instead characteristic of ancient ancestry in a geographic region where malaria was endemic. Africa is one such region, but so are the Mediterranean and southern India. If sickle-cell disease is suspected, then the correct diagnostic approach is not simply to determine the patient's race, but to ask whether they have African, Mediterranean or South Indian ancestry. To use genotype effectively in making diagnostic and therapeutic decisions, it is not race that is relevant, but both intra- and trans-continental contributions to a person's ancestry.

Race and ancestry are confounded both by genetic heterogeneity within groups and by the widespread mixing of previously isolated populations. The assignment of a racial classification to an individual hides the biological information that is needed for intelligent therapeutic and diagnostic decisions. A person classified as 'black' or 'Hispanic' by social convention could have any mixture of ancestries, as defined by continent of origin. Confusing race and ancestry could be potentially devastating for medical practice.

Other attempts to classify people into broad genetic groups based on the frequency of specific genes for, say, drug-metabolizing enzymes, are also likely to be poor predictors of medical outcome. As with racial groupings, the overall variation in the frequencies of such genes between groups is likely to be less than that within each group.

The conventional, social definition of race is useful in a medical context as it provides information about the social circumstances and lifestyle of patients. But this is a consequence of social history, so any variation is (at least in principle) transitory. By contrast, information on the likelihood that a person carries specific disease-related or treatment-response genes is grounded in their ancestry in far more complex ways. We suggest that identifying all contributions to a patient's ancestry can be useful in diagnosing and treating diseases with genetic influences. Eventually, for both diagnosis and treatment, specific genetic variants will provide concrete, useful information. ■

Marcus W. Feldman is in the Department of Biological Sciences, Stanford University, California 94305, USA.
Richard C. Lewontin is at the Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA.
Mary-Claire King is in the Departments of Genome Science and Medicine, University of Washington, Seattle, Washington 98195, USA.

Race

Ancestral genetic data are far more useful for medical purposes than are racial categories, which may be correlated with disease for social or economic rather than biological reasons.

FURTHER READING

Rosenberg, N. A. *et al. Science* **298**, 2381–2385 (2002).
González Burchard, E. *et al. N. Engl. J. Med.* **348**, 1170–1175 (2003).
Cooper, R. S., Kaufman, J. S. & Ward, R. *N. Engl. J. Med.* **348**, 1166–1170 (2003).
Lewontin, R. C. *Evol. Biol.* **6**, 381–398 (1972).
Barbujani, G., Magagni, A., Minch, E. & Cavalli-Sforza, L. L. *Proc. Natl Acad. Sci. USA* **94**, 4516–4519 (1997).

62

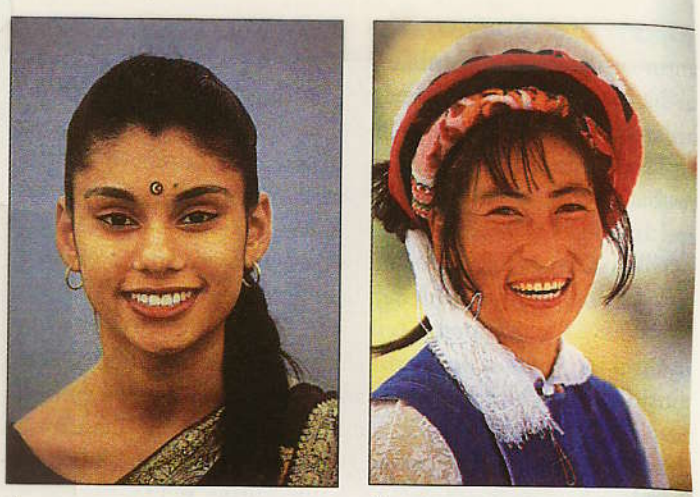
HUMAN DIVERSITY

RICHARD LEWONTIN

ARE There
HUMAN Races
or
Human Beings?

DNA testing CAN provide information on Human origins - geographical groupings & Ancestry!

We will Learn all about HUMAN VARIATION FROM GENOME Sequencing!



A. B.

Figure 14-4 **People from different geographic regions vary in such traits as height, blood type, skin color, and facial features.** A. An Indian woman. B. An ethnic minority woman from China. *(A, Superstock; B, Paul Grebliunas/Tony Stone Images)*

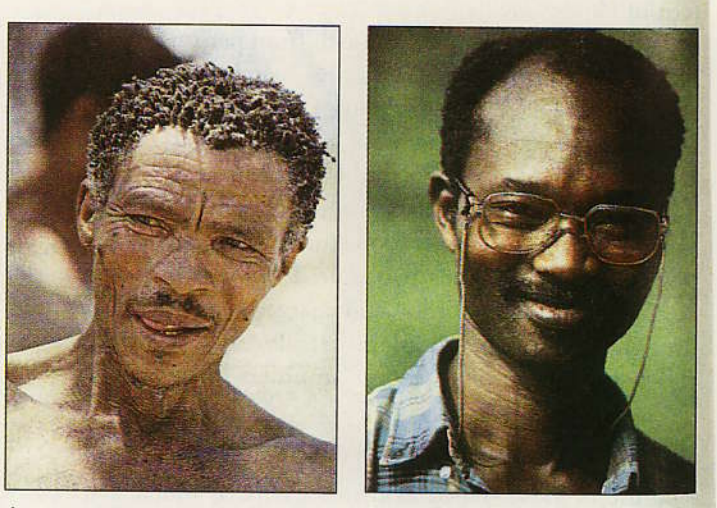
What Are Human "Races"?

We all know that the human species varies in appearance and in physiology geographically. People of Indian descent, for example, are recognizably different from those of Chinese descent. We can look at the shape of the nose, the eyes, and the ears, for example, and we see differences (Figure 14-4). We can look at the distribution of individual alleles and see differences. Fifty-one percent of Nigerians have type O blood, compared to only 30 percent of Japanese. Twenty percent of Russians have type B blood, while the Amerindians of Lima, Peru, have no detectable levels of type B blood at all.

Nineteenth-century anthropologists struggled to classify human groups into a few major races. Some systems identified only 12 races, while other systems listed 30 or more. One problem was that no matter how anthropologists classified humans, there always seemed to be tribes or nations that would not fit into any known group. The Basques, who live in the Pyrenees mountains between France and Spain, for example, appear European. Yet their language and culture are unlike any other in the world, and some researchers once argued that they are direct descendants of Stone Age Europeans. Similarly, the Bushmen are unique among African groups in both appearance and physiology (Figure 14-5).

A more serious problem with the grouping of humans into races is that most groups do *not* stand out from those around them; they blend. Because groups of humans inevitably mix, through migration, warfare, and trade, human "races" are never pure. Both the Japanese and British, for example, take pride in the purity of their island races. Yet the Japanese are a grade mixture of Korean and Ainu north islanders (a people of possibly European descent). This mix shows up in the distribution of blood types from one end of Japan to the other (Figure 14-6).

The British are even more of a melting pot than the Japanese. The Bronze Age Beaker Folk mixed with the Indo-European Celts in the first thousand years B.C. In the next thousand years, the Angles, the Saxons, the Jutes, and the Picts arrived, followed by the Vikings and their descendants, the



A. B.

Figure 14-5 **Classifying humans into a few discrete races has been unsuccessful.** Named races frequently include markedly distinct peoples. A. A Bushman from Namibia. B. A rubber plantation foreman from the Ivory Coast. *(A, M.P. Kahl/Photo Researchers, Inc.; B, Charles O. Cecil/Visuals Unlimited)*

SIMILARITY AND DIFFERENCES in Blood Group Allele Frequencies by "Race" or Population

How does Genetic Variation (different alleles) VARY Between & within Populations?

Examples of extreme differentiation and close similarity in blood group allele frequencies in three racial groups

(1) Different variation but most Alleles Present →

(2) SAME VARIATION →

Gene	Alleles	Caucasoid	Negroid	Mongoloid
Duffy	*Fy*	.0300	.9393	.0985
	Fy^a	.4208	.0607	.9015
	Fy^b	.5492	—	—
Rhesus	R_0	.0186	.7395	.0409
	R_1	.4036	.0256	.7591
	R_2	.1670	.0427	.1951
	r	.3820	.1184	.0049
	r'	.0049	.0707	0
	others	.0239	.0021	0
P	P_1	.5161	.8911	.1677
	P_2	.4839	.1089	.8323
Auberger	Au^a	.6213	.6419	
	Au	.3787	.3581	
Xg	Xg^a	.67	.55	.54
	Xg	.33	.45	.46
Secretor	*Se*	.5233	.5727	
	se	.4767	.4273	

Source: R. C. Lewontin, *The Genetic Basis of Evolutionary Change* (Columbia University Press, 1974).

(1) Most alleles in ALL "races" / populations

(2) NO Homozygousity at any locus

(3) Some Allelic Frequencies differ Between "Races" & Some are the Same! Duffy differs & Xg is the SAME. Why? Adaptive value.

(4) Auberger, Xg, & Secretor Loci show have alleles vary within populations similarly & show NO between population differences.

(5) Wide range & Different Alleles within / Between "Races"

Similarity & Differences in Blood Group Gene Alleles

VARIATION Between Groups

26-2 TABLE Examples of Extreme Differentiation and Close Similarity in Blood Group Allelic Frequencies in Three Racial Groups

		POPULATION		
Gene	Allele	Caucasoid	Negroid	Mongoloid
Duffy	*Fy*	0.0300	0.9393	0.0985
	Fy^a	0.4208	0.0000	0.9015
	Fy^b	0.5492	0.0607	0.0000
Rhesus	R_0	0.0186	0.7395	0.0409
	R_1	0.4036	0.0256	0.7591
	R_2	0.1670	0.0427	0.1951
	r	0.3820	0.1184	0.0049
	r′	0.0049	0.0707	0.0000
	Others	0.0239	0.0021	0.0000
P	P_1	0.5161	0.8911	0.1677
	P_2	0.4839	0.1089	0.8323
Auberger	Au^a	0.6213	0.6419	No data
	Au	0.3787	0.3581	No data
Xg	Xg^a	0.67	0.55	0.54
	Xg	0.33	0.45	0.46
Secretor	*Se*	0.5233	0.5727	No data
	se	0.4767	0.4273	No data

Source: A summary is provided in L. L. Cavalli-Sforza and W. F. Bodmer, *The Genetics of Human Populations* (W. H. Freeman and Company, 1971), pp. 724–731. See L. L. Cavalli-Sforza, P. Menozzi, and A. Piazza, *The History and Geography of Human Genes* (Princeton University Press, 1994), for detailed data.

Wide VARIATION

Wide VARIATION

Little VARIATION

We all Have the SAME Genes But MANY differ in Frequencies of SOME alleles / genes / SNPs

Differences in Allelic Frequencies in various populations

24-5 TABLE Frequencies of the Alleles I^A, I^B, and i at the ABO Blood Group Locus in Various Human Populations

Population	I^A	I^B	i
Eskimo	0.333	0.026	0.641
Sioux	0.035	0.010	0.955
Belgian	0.257	0.058	0.684
Japanese	0.279	0.172	0.549
Pygmy	0.227	0.219	0.554

Source: W. C. Boyd, *Genetics and the Races of Man.* D. C. Heath, 1950.

24-1 TABLE Frequencies of Genotypes for Alleles at MN Blood Group Locus in Various Human Populations

	GENOTYPE			ALLELE FREQUENCIES	
Population	*M/M*	*M/N*	*N/N*	*p(M)*	*q(N)*
Eskimo	0.835	0.156	0.009	0.913	0.087
Australian aborigine	0.024	0.304	0.672	0.176	0.824
Egyptian	0.278	0.489	0.233	0.523	0.477
German	0.297	0.507	0.196	0.550	0.450
Chinese	0.332	0.486	0.182	0.575	0.425
Nigerian	0.301	0.495	0.204	0.548	0.452

Source: W. C. Boyd, *Genetics and the Races of Man.* D. C. Heath, 1950.

24-2 TABLE Frequencies of Gametic Types for MNS System in Various Human Populations

	GAMETIC TYPE				HETEROZYGOSITY (*H*)	
Population	*M S*	*M s*	*N S*	*N s*	From gametes	From alleles
Ainu	0.024	0.381	0.247	0.348	0.672	0.438
Ugandan	0.134	0.357	0.071	0.438	0.658	0.412
Pakistani	0.177	0.405	0.127	0.291	0.704	0.455
English	0.247	0.283	0.080	0.290	0.700	0.469
Navaho	0.185	0.702	0.062	0.051	0.467	0.286

Source: A. E. Mourant, *The Distribution of the Human Blood Groups.* Blackwell Scientific, 1954.

Reasons for Allelic variation between Populations

1. **Founder Effect / Geographical Isolation**
   - ↳ Selective mating due to geography / culture or both

2. **Adaptive Value**
   - ↳ Have positive effect in specific environments
     - e.g.: HB^S Sickle-Cell Globin allele (India, Africa, Mediterranean)
     - Duffy
     - Skin Color Genes

GEOGRAPHY &/or ADAPTIVE VALUE

But do not vary across whole genome as most loci are "neutral"

NOVEL EXAMPLE OF ALLELE ADAPTIVE VALUE

(1) $Fy^{B^{ES}}$ (erythroid silent) is a Major Duffy allele in African Americans & Black populations of African heritage. Rarely in other populations!

(2) Encodes chemokine-receptor protein on blood cell membrane. Parasite receptors for Malarial PLASMODIUM bind to this Receptor.

(3) $Fy^{B^{ES}}$ is a <u>switch</u> mutation allele — $Fy^{B^{ES}}$ allele <u>CANNOT</u> be TRANSCRIBED ∴ No Fy^B or PLASMODIUM receptor protein —

(4) NO binding of Malaria PLASMODIUM — ∴ Malarial RESISTANCE

(5) High FREQUENCY in Geographical Regions with high Levels of Malaria!

(6) Other Duffy alleles (CHROMOSOME 1) @ different frequencies in other populations

There is A Large variation in D1S80 VNTR alleles within populations but Little between

Table 15.2 Allele frequencies for *D1S80* among U.S. population groups

Repeat number	Caucasian	Hispanic	African American	Asian
14	0	0	0	0
15	0	0.001	0	0
16	0.001	0.010	0.002	0.034
17	0.002	0.009	0.028	0.025
18	0.237	0.224	0.073	0.152
19	0.003	0.005	0.003	0.022
20	0.018	0.013	0.032	0.007
21	0.021	0.028	0.115	0.034
22	0.038	0.024	0.081	0.017
23	0.012	0.009	0.014	0.017
24	0.378	0.315	0.234	0.230
25	0.046	0.072	0.045	0.027
26	0.020	0.007	0.006	0
27	0.007	0.016	0.008	0.047
28	0.063	0.078	0.130	0.076
29	0.052	0.055	0.053	0.042
30	0.008	0.039	0.009	0.123
31	0.072	0.053	0.054	0.093
32	0.006	0.005	0.007	0.012
33	0.003	0.004	0.004	0.005
34	0.001	0.006	0.086	0.005
35	0.003	0	0.002	0.005
36	0.004	0.011	0.001	0.005
37	0.001	0.004	0	0.007
38	℘	℘	℘	℘
39	0.003	0.004	0.003	0.005
40	0	0	0	0
41	0	0.002	0.002	0.007
>41	0.001	0.006	0.007	0.002
Sample size	718	409	606	204

Source: Data from B. Budowle, et al. 1995. *Journal of Forensic Science* 40:38

JUST AS Predicted FROM Class Genotype

VNTR used for HC70A DNA Finger print

NO ADAPTIVE VALUE!! Good LOCUS For Forensics - Group Neutral!

There is More Genetic Diversity within Populations than Between Populations!! So much for the concept of racial "purity"!!!

Proportion of genetic diversity accounted for within and between populations and races

Gene	Total $H_{species}$	Proportion: Within Populations	Within Races between Populations	Between Races
Hp	.994	.893	.051	.056
Ag	.994	.834	—	—
Lp	.639	.939	—	—
Xm	.869	.997	—	—
Ap	.989	.927	.062	.011
6PGD	.327	.875	.058	.067
PGM	.758	.942	.033	.025
Ak	.184	.848	.021	.131
Kidd	.977	.741	.211	.048
Duffy	.938	.636	.105	.259
Lewis	.994	.966	.032	.002
Kell	.189	.901	.073	.026
Lutheran	.153	.694	.214	.092
P	1.000	.949	.029	.022
MNS	1.746	.911	.041	.048
Rh	1.900	.674	.073	.253
ABO	1.241	.907	.063	.030
Mean		.854	.083	.063

Source: R. C. Lewontin, *Genetic Basis of Evolutionary Change* (Columbia University Press, 1974).

More genetic diversity within any population than between populations!

within any population

of same Race

(1) 85% of Human Genetic Variation occurs within Populations & between Individuals in that population!

(2) Remaining 15% of Human Genetic Variation Split between different populations of same "Race" (8%) & between different "Races" (6%).

(3) Only 6% of Human Genetic Variation due to Differences between Races!! Geographic

The SAME ALLeLes are present in MOST Populations But at different Frequencies

Mapping Human History

Mary-Claire King and Arno G. Motulsky

The DNA of modern humans contains a record of the travels and encounters of our ancestors. The genotypes of people living today are the result of ancient human migrations, the continuous appearance of new mutations, selection by climate and infection for genetic alleles that conferred a survival advantage, and mating patterns determined by cultural norms. By sampling genotypes from people across the globe, geneticists have reconstructed the major features of our history: our ancient African origin, migrations out of Africa, movements and settlements throughout Eurasia and Oceania, and peopling of the Americas (*1*–*5*). As genomic technology has improved, these analyses of genotype have successively incorporated new markers: blood groups (*2*), protein polymorphisms (*2*), mitochondrial DNA sequences (*1*), Y chromosome haplotypes (*3*), and highly variable nuclear microsatellite markers (*4, 5*).

The most recent contribution to this literature is by Rosenberg *et al.* (*6*) on page 2381 of this issue. These investigators explored the genetic structure of human populations using highly variable markers on the human autosomes of individuals from different parts of the world. The genotyped markers were microsatellite short tandem repeat sequences that do not encode any expressed genes and are generally selectively neutral. The populations studied were defined by geography, language, and culture, and participating individuals were well rooted in their populations, with several generations of ancestors known to have lived in the same locale as the participant. Genotypes from more than a thousand individuals were evaluated by a statistical method that defines clusters of people on the basis of genetic similarity at multiple loci, without using prior information about ancestry. In this method, individuals are assigned to clusters probabilistically (*5, 7*). Individuals may have significant probabilities of membership in more than one cluster due either to genetic similarities of groups or to ancestral intergroup matings. The world map (see the figure) illustrates variation at one microsatellite marker in 12 populations. This marker has four common alleles, each of which appears in all populations. Rare alleles are shared by fewer populations. Few alleles are unique to only one population. No allele is population specific.

Previous genetic analyses of human history have consistently suggested that most human genetic variation is due to differences among individuals within populations rather than to differences among populations (*4, 8*). The Rosenberg *et al.* analysis of many more markers and many more people confirms this result: 93 to 95% of genetic variation is due to genetic differences among individuals who are members of the same population and only 3 to 5% of genetic variation is due to differences among the major population groups.

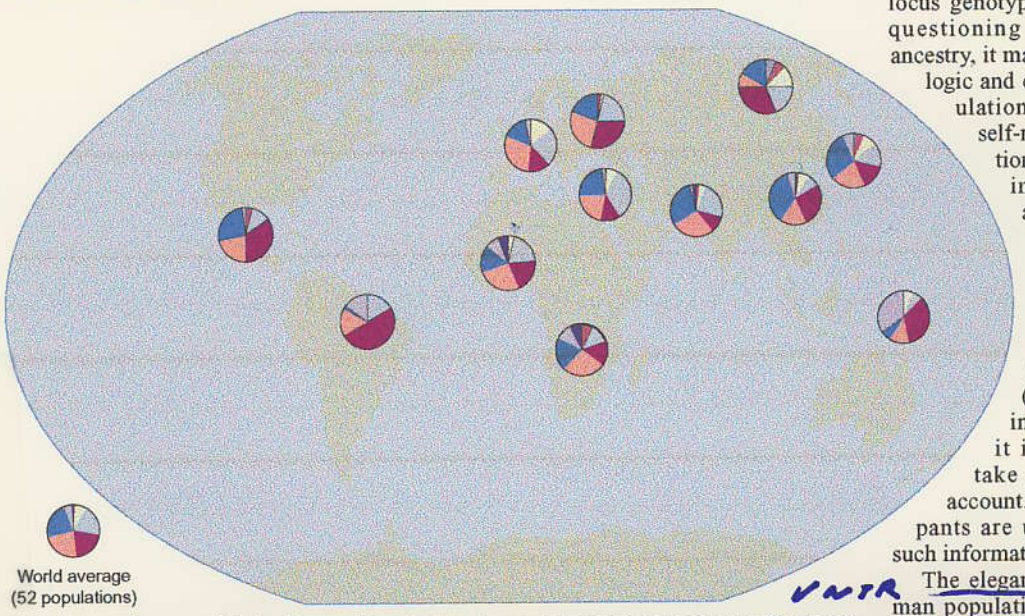
The power of the method lies in the construction of clusters on the basis of accumulated small differences in allele frequencies across many markers and many people. Statistical clustering of genotypes—composed of 4682 alleles from 377 markers in 1056 individuals from 52 populations—yields groups corresponding to major geographic regions of the world [see figure 1 in (*6*)]. Creation of two clusters reflects ancient human origins in Africa and rapid expansion throughout Eurasia, and migrations to the Americas from East Asia. Creation of five clusters yields groups corresponding to five major geographic regions of the world: Africa, Eurasia (Europe, the Middle East, Central and South Asia), East Asia, Oceania, and America. There is excellent agreement between membership of individuals in these clusters and their self-identified regions of origin. Similar results were obtained by the same statistical approach based on fewer populations and fewer markers [Table 2 of (*5*)].

The authors are in the Department of Medicine (Medical Genetics) and Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA. E-mail: mcking@u.washington.edu

We Learn about our origins + diversity from the Human Genome Project

12

Population substructure could be consistently identified within some geographic regions but not others. Within Africa, for example, analysis consistently yielded the same four subclusters: Mbuti Pygmies, Biaka Pygmies, San peoples, and speakers of Niger-Kordofanian languages (Bantu, Yoruba, and Mandenka populations). In contrast, within Europe, multiple analyses were not consistent. Many more individuals will need to be included to sort out European demographic history.



Humans on the move. Worldwide genetic variation at a neutral marker. Allele frequencies of one randomly chosen microsatellite marker reveal common alleles shared in all populations and the gradual and arbitrary differences in allele frequencies across geographic regions. Populations shown in this example are Yoruba and Bantu (Africa); French, Russians, Palestinians, and Pakistani Brahui (Eurasia); Han Chinese, Japanese, and Yakut (East Asia); New Guineans (Oceania); and Maya and Karitianans (America). Each color on the diagrams represents one of nine alleles of GGAA29H03 (D13S1493), which range in length from 219 to 255 base pairs. By accumulating small differences in allele frequencies from hundreds of such highly variable markers and hundreds of people, statistical methods reveal genetic clusters of Africans, Eurasians, East Asians, Pacific Islanders, and Americans, corresponding to major ancient human migrations (*6*).

The identification of clusters corresponding to the major geographic regions may depend on the sampling of individuals from well-defined, relatively homogeneous populations. If individuals were sampled from a worldwide "grid" (or a worldwide grid weighted by population density), the clusters might be much less precisely defined. Does the correspondence of worldwide genetic clusters and major geographic regions suggest borders around genetic clusters analogous to the physical borders—oceans, mountain ranges, and deserts—separating geographic regions? No. Both the results of Rosenberg and colleagues and those of previous studies (*1*–*5*, *8*) indicate that unlike separations between geographic regions, differences in allele frequencies are gradual, without discontinuities between clusters. After thousands of years—if enough markers and people are studied—allele frequency differences are collectively adequate to create clusters that correspond to the major migrations of human history.

What are the implications of the Rosenberg *et al.* findings for medicine? The current medical literature increasingly includes studies exploring population differences in disease incidence or in efficacy or adverse responses to drug treatment (*5*, *9*). The rationale of these studies is that alleles influencing disease susceptibility or treatment response may differ in frequency across populations. Consequently, individuals would be better served if critical genotypes were taken into account when assessing disease risk or designing treatment regimens. In the absence of knowing the identities of the critical alleles, personal ancestry as indicated by study participants is often used as an initial but potentially misleading substitute.

The Rosenberg *et al.* data suggest that with the exception of ancient highly selected loci (for example, the Duffy null blood group, which confers complete protection against vivax malaria), very few alleles will be both confined to one population and common enough in that population to dictate medical management for the entire group. Instead, critical alleles influencing disease risk or response to treatment are likely to be either ancient, worldwide, and relatively common in many populations, or geographically localized and individually rare (*10*).

Differences among populations in disease frequency and treatment outcome certainly occur but may not be genetic in origin. Given that the major population origin of groups can be defined by multilocus genotype clustering (*5*–*7*) without questioning individuals about their ancestry, it may be tempting in epidemiologic and clinical studies to omit population characterization through self-reporting. However, correlations identified by the clustering method may be falsely ascribed to genes when in fact they had nothing to do with genetics but were caused by social, economic, or discriminatory factors limited to a genetically defined population cluster (*11*). To evaluate medically important group differences, it is therefore necessary to take all such risk factors into account. Patients and study participants are usually the best source of such information.

The elegant statistical analysis of human population structure by Rosenberg and colleagues reflects the major human migrations out of Africa, into Europe, across Asia, into Oceania, and to the Americas. By genotyping a large sample of an individual's alleles, it is possible to identify the migrations in which his or her ancestors participated. But the link between historical genetic demography and medically important risk is complex. Disease susceptibility may be genetic but not geographically clustered, or geographically clustered but not genetic, or neither, or both.

References

1. A. C. Wilson, R. L. Cann, *Sci. Am.* **266**, 68 (April 1992).
2. L. L. Cavalli-Sforza, P. Menozzi, A. Piazza, *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, NJ, 1994).
3. P. A. Underhill *et al.*, *Nature Genet.* **26**, 358 (2000).
4. G. Barbujani, A. Magagni, E. Minch, L. L. Cavalli-Sforza, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4516 (1997).
5. J. F. Wilson *et al. Nature Genet.* **29**, 265 (2001).
6. N. A. Rosenberg *et al. Science* **298**, 2381 (2002).
7. J. K. Pritchard, M. Stephens, P. Donnelly, *Genetics* **155**, 945 (2000).
8. R. C. Lewontin, *Evol. Biol.* **6**, 381 (1972).
9. H. G. Xie, R. B. Kim, A. J. Wood, C. M. Stein, *Annu. Rev. Pharmacol. Toxicol.* **41**, 815 (2001).
10. J. K. Pritchard, N. J. Cox, *Hum. Mol. Genet.* **11**, 2417 (2002).
11. N. Risch, E. Burchard, E. Ziv, H. Tang, *Genome Biol.* **3**, comment2007 (2002).

CONCLUSION

(1) If 85% of Human Genetic Variation occurs between different People within any given population (localized)

(2) If only 7% of Human Genetic Variation occurs Between "RACES" (Novel ALLELES Specific to a "race") → e.g., $Fy\alpha^{ES}$

(3) Then Losing all "races" Except one retains 94% of all Human Genetic Variation!

$[85\% + (15\% - 7\%)] = 94\%$

VARIATION That Occurred in ANCESTRAL Populations

85% within population genetic variability
8% between populations of SAME "race"
7% between "race" genetic variability

(4) ∴ Humans highly Heterozygous or Hybrids —
* if above not TRUE - most of us would not be here — need genetic variation to survive!

So what is a "Race"?

(1) Primarily a sociological concept — but could be a localized or inbred population that has a higher frequency of alleles at a very small number of loci. Affects few physical features.

(2) High frequency alleles in one "race" are present at lower frequency in other "Races." ALL Humans have SAME Genes — differ in form Mostly within populations!

(3) Heterozygosity (Variation) high in human populations — ALL populations. None homozygous at all Loci!

(4) NO such thing as a "pure" Race — would have little variation —

(5) Genes Affecting Physical features NOT representative of genes across genome —

Geographical Ancestry is Relevant — Many "racial" groups may have Multiple Ancestries because of Admixture & Migration

∴ "Race" Classification is Arbitrary & Based on a few Traits — Not Science Based

CAN DEFINE BY MANY Criteria ...

RACE BY RESISTANCE

Traditionally we divide ourselves into races by the twin criteria of geographic location and visible physical characteristics. But we could make an equally reasonable and arbitrary division by the presence or absence of a gene, such as the sickle-cell gene, that confers resistance to malaria. By this reckoning, we'd place Yemenites, Greeks, New Guineans, Thai, and Dinkas in one "race," Norwegians and several black African peoples in another.

People who posses anti-Malarial genes (alleles!) & those that do not

RACE BY DIGESTION

We could define a race by any geographically variable trait—for example, the retention in adulthood of the enzyme lactase, which allows us to digest milk. Using this as our divisive criterion, we can place northern and central Europeans with Arabians and such West African peoples as the Fulani; in a "lactase-negative race," we can group most other African blacks with east Asians, American Indians, southern Europeans, and Australian aborigines.

People who possess lactase & those that do not!

RACE BY FINGERPRINTS

Probably the most trivial division of humans we could manage would be based on fingerprint patterns. As it turns out, the prevalence of certain basic features varies predictably among peoples: in the "Loops" race we could group together most Europeans, black Africans, and east Asians. Among the "Whorls" we could place Mongolians and Australian aborigines. Finally, in an "Arches" race, we could group Khoisans and some central Europeans.

People who possess Loops & those that have arches!

RACE BY GENES

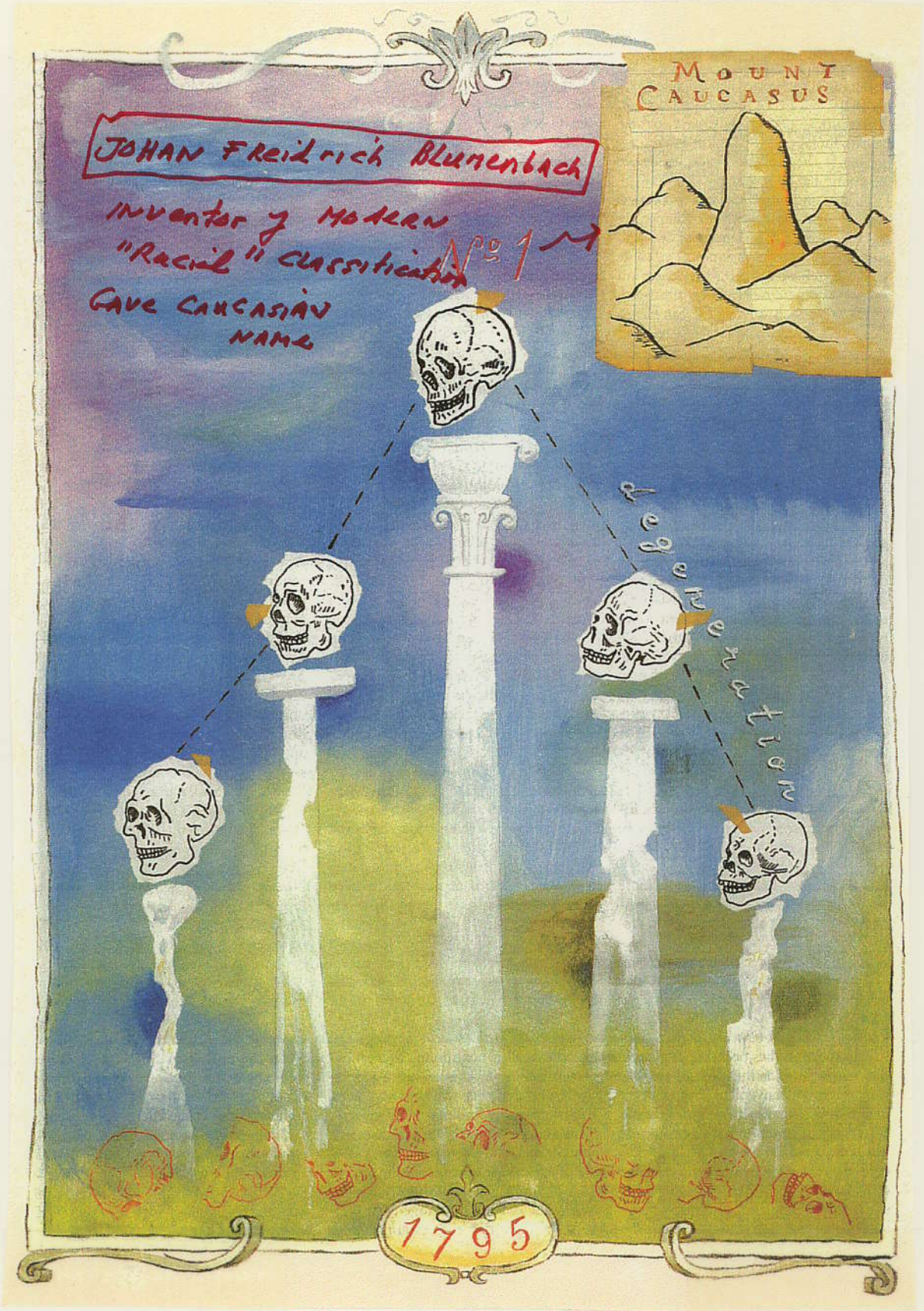
One method that seems to offer a way out of arbitrariness is to classify peoples by degree of genetic distinctness. By this standard the Khoisans of southern Africa would be in a race by themselves. African blacks would form several other distinct races. All the rest of the world's peoples—Norwegians, Navajo, Greeks, Japanese, Australian aborigines, and so on—would, despite their greatly differing external appearance, belong to a single race.

People who possess one form of an allele (localized inbred population) & those that do not

RACIAL CLASSIFICATION DIDN'T COME FROM SCIENCE BUT FROM THE BODY'S SIGNALS FOR DIFFERENTIATING ATTRACTIVE FROM UNATTRACTIVE SEX PARTNERS, AND FRIEND FROM FOE.

Physical traits

HOW DID WE GET TO WHERE WE ARE?



JOHAN FREIDRICH BLUMENBACH

(1) Changed Linnaean Neutral-geography-based Human Classifications to a value-based classification—

(2) Said all "races" originated from one place/origin – in Europe around Mt. CAUCASUS – because they are the most "beautiful" race – even though he felt all races were equal in all respects & argued with people that didn't!

(3) Classified five "races"

- CAUCASIAN
- Mongolian
- Ethiopian
- American
- Malay

Value-Based Classification & putting Caucasians on top – most beautiful – had disasterous consequences!

CONCLUDE

(1) Races are Arbitrary Entities - Social Constructs that are Culturally generated.

(2) Yes - there is genetic differences between "races" or relatively inbred geographical populations that can lead to physical & other differences - due to very small # genes & not reflective of whole genome

(3) Within population genetic variation Much greater than between population genetic variation - Many loci have same allele frequencies - some differ

(4) Only minor differences between genomes of different people or groups of people - Unity >>> differences!

(5) We are all the same - but different.

you now know why!

Race classifications Arbitrary, unscientific, & divisive!


The End!!